Exploring the diversity of CRESS DNA viruses associated with the faecal matter of wild and domestic animals in New Zealand

---

A thesis submitted in partial fulfilment of the requirements for the

Degree of Master of Science in Microbiology

at the University of Canterbury

By Olivia Steel

University of Canterbury

2016

Table of Contents

List of Tables

Chapter One: Literature Review

Chapter Two: Isolation and characterisation of CRESS DNA viruses and circular molecules from animal faecal matter

Chapter Three: Concluding remarks

List of Figures

Chapter One: Literature Review

Chapter Two: Isolation and characterisation of CRESS DNA viruses and circular molecules from animal faecal matter

Acknowledgements

This thesis would not have been possible without the support and guidance of the virology lab group, particularly Dr. Arvind Varsani and Dr. Simona Kraberger. Thank you for accepting me into your lab group and helping me whenever I had issues, no matter how small. I am very grateful for your patience and for all that you have taught me over the past year. It has been such a pleasure and a privilege to work with you all.

I would also like to acknowledge the staff members and students at the University of Canterbury who collected the animal faecal samples used in my research. Thank you Aaron Stevens, Alyssa Sikorski, Anisha Dayaram, Daisy Stainton, Jenny Ladley, Laura Young, Laurel Julian, Ryan Catchpole, Simona Kraberger and Arvind Varsani.

Finally, I would like to thank my friends and family. Thank you for all your patience and encouragement throughout this stressful time. You never failed to cheer me up when things just weren't going right. Thank you for always being there whenever I needed you, even if it was just for a quick pep talk or a much-needed coffee.

Abstract

Before the advent of cost-effective, efficient next-generation sequencing technologies, viral research predominantly focussed on viruses that were easily identified and those of importance to humans i.e. were pathogenic. Advances in viral metagenomic approaches in recent years have revolutionised the discovery of novel viruses. In particular, the number of described single-stranded DNA (ssDNA) viruses has grown rapidly. The increased rate of discovery revealed that ssDNA viruses a have higher abundance and diversity than previously thought. An increasing number of ssDNA viruses identified through metagenomic studies are too divergent to be classified into the current taxonomic system established by the International Committee on Taxonomy of Viruses and remain unclassified.

Circular replication-associated protein (Rep)-encoding single-stranded (CRESS) DNA viruses have been identified in a variety of environmental samples using viral metagenomic approaches. A high number of CRESS DNA viruses are unclassified, prompting the proposal of new groupings, namely the gemycircularviruses, smacoviruses and cycloviruses. The studies that identified CRESS DNA viruses were largely conducted outside New Zealand up until 2013. New Zealand's Gondwana ancestry and physical isolation provides a unique virome that has not been extensively explored. Further exploration of CRESS DNA viruses within New Zealand would shed light on the true diversity and prevalence of CRESS DNA viruses both within New Zealand and globally, and may provide support to proposed CRESS DNA virus groupings.

Faecal sources harbour many potential host species and enable sampling of the virome of an ecosystem in a non-invasive manner. Accordingly, faecal sources have been used effectively to discover a wide variety of viruses in multiple studies. This dissertation utilised a viral metagenomic approach to identify CRESS DNA viruses in the faeces of wild and domestic animals sampled from sites across the South Island of New Zealand, expanding upon baseline data collected by other studies. Next-generation sequencing technologies were used to inform the design of specific

abutting primers for the recovery of complete viral genomes from faecal matter. This approach recovered 38 complete CRESS DNA viral genomes and two circular molecules from 49 individual faecal samples. The recovered viruses were classified as gemycircularviruses (n=18), smacoviruses (n=12) or unclassified CRESS DNA viruses (n=8) according to BLASTx similarities in GenBank's non-redundant database, genome-wide nucleotide pairwise identities and shared genome organisations with previously identified CRESS DNA viruses. The eighteen gemycircularvirus isolates represent eleven species, nine of which are novel, while the twelve smacoviruses constitute eleven species, including ten novel species. The remaining CRESS DNA viruses could not be classified and represent unique species. The CRESS DNA viruses identified in this study show a wide diversity and contribute significantly to our understanding of the prevalence and diversity of CRESS DNA viruses circulating within New Zealand.

Chapter One: Literature Review

Table of Contents

1.1 General introduction to viruses

Viral hosts span all domains of life- Archaea, Bacteria and Eukarya (King *et al.*, 2011). Accordingly, viruses are found in a wide range of ecosystems, including many extreme environments (Le Romancer *et al.*, 2007). Studies have identified viruses from the Namib, Sahara, Death Valley and Kutch deserts (Adriaenssens *et al.*, 2015; Fancello *et al.*, 2013; Kerepesi & Grolmusz, 2015; Prestel *et al.*, 2013; Prigent *et al.*, 2005), numerous hypersaline and alkaline environments (Atanasova *et al.*, 2012; Emerson *et al.*, 2012; Jiang *et al.*, 2004; Oren *et al.*, 1997; Roine *et al.*, 2010; Santos *et al.*, 2011; Sime-Ngando *et al.*, 2011), deep-sea thermal vents (Geslin *et al.*, 2003; Ortmann & Suttle, 2005; Williamson *et al.*, 2008), hot springs (Bize *et al.*, 2008; Bolduc *et al.*, 2012; Breitbart *et al.*, 2004; Häring *et al.*, 2005; Rachel *et al.*, 2002), and Antarctic ecosystems (Kepner *et al.*, 1998; Laybourn-Parry *et al.*, 2001; López-Bueno *et al.*, 2009; Zablocki *et al.*, 2014; Zawar-Reza *et al.*, 2014).

Developments in molecular techniques and next-generation sequencing technologies enabled a rapid rate of discovery of divergent and novel viruses from a range of sample types (Breitbart *et al.*, 2003; Cantalupo *et al.*, 2011; Kim *et al.*, 2008; Ng *et al.*, 2011a; Rosario *et al.*, 2011; Rosario *et al.*, 2009b; Roux *et al.*, 2012; Whon *et al.*, 2012). The rate of discovery suggests that the size and diversity of the global virome is likely underestimated.

The Baltimore classification system groups viruses into one of seven groups according to the nature of their genome: double-stranded DNA, single-stranded DNA, reverse transcribing viruses with either DNA or RNA genomes, double-stranded RNA, negative-sense single-stranded RNA, and positive-sense single-stranded RNA (Baltimore, 1971). The International Committee on Taxonomy of Viruses (ICTV) is responsible for developing a taxonomic scheme for the classification of novel viruses and subviral agents (King *et al.*, 2011). Isolates are assigned to taxa according to sequence similarity, genome organisation, replication mechanism, morphology of particles and various biological properties (King *et al.*, 2011). The ICTV determined that a viral species is "a polythetic class of viruses that constitutes a replicating lineage and occupies a particular ecological niche" (King *et al.*, 2011). The large

number of unclassified viruses indicates that the current view of viral taxonomy is incomplete.

1.2 Introduction to single-stranded DNA viruses

1.2.1 Classification of single-stranded DNA viruses

Technological advancements have contributed to the rapid rate of discovery of single-stranded DNA (ssDNA) viruses. The diverse nature of unclassified ssDNA viruses indicates that the current classification scheme employed by the ICTV may require revision (Rosario *et al.*, 2012b). Currently, ssDNA viruses are divided into nine recognised families: *Anelloviridae, Bidnaviridae, Circoviridae, Geminiviridae, Inoviridae, Microviridae, Nanoviridae, Parvoviridae* and *Spiraviridae* (Table 1.1) (Hu *et al.*, 2013; King *et al.*, 2011; Mochizuki *et al.*, 2012).

Table 1.1: Characteristics of ssDNA viral families

Virus Family	Host	Morphology	Genome Configuration	Genome Size
Anelloviridae	Vertebrates	Icosahedral	Circular ssDNA, monopartite	2.0-3.9kb
Bidnaviridae	Invertebrates	Icosahedral	Linear ssDNA, bipartite	6.5kb and 6.0kb
Circoviridae	Vertebrates	Icosahedral	Circular ssDNA, monopartite	1.7-2.0kb
Geminiviridae	Plants	Geminate	Circular ssDNA, monopartite/bipartite	2.5-3.0kb
Inoviridae	Bacteria	Filamentous/ Rod-shaped	Circular ssDNA, monopartite	6.0-12.4kb
Microviridae	Bacteria	Icosahedral	Circular ssDNA, monopartite	5.3-6.1kb
Nanoviridae	Plants	Icosahedral	Circular ssDNA, multipartite	0.9-1.1kb
Parvoviridae	Vertebrates/ Invertebrates	Icosahedral	Linear ssDNA, monopartite	4.0-6.3kb
Spiraviridae	Archaea	Cylindrical helix	Circular ssDNA, monopartite	24.9kb

1.2.1.1 Animal-infecting ssDNA viruses

Animal-infecting viruses are classified into the *Anelloviridae, Circoviridae Parvoviridae* and *Bidnaviridae* families that infect varied hosts. Anelloviruses infect multiple hosts including primates, rodents, dogs and pigs (King *et al.*, 2011; Leary *et al.*, 1999; Mushahwar *et al.*, 1999; Nishiyama *et al.*, 2014; Okamoto *et al.*, 2002). Circoviruses are known pathogens of pigs and birds, although isolates have been recovered from human faeces, dogs, foxes, bats, amphibians and fish samples (Allan *et al.*, 1998; Blinkova *et al.*, 2010; Ellis *et al.*, 1998; Kapoor *et al.*, 2012; King *et al.*, 2011; Li *et al.*, 2010a; Li *et al.*, 2010b; Lorincz *et al.*, 2011; Ritchie *et al.*, 1989; Sauvage *et al.*, 2011; Tarján *et al.*, 2014; Tischer *et al.*, 1982; Todd *et al.*, 1990).

Members of the *Parvoviridae* family infect vertebrates (*Parvovirinae* subfamily*)* and invertebrates (*Densovirinae* subfamily), including humans, cows, shrimp and mosquitoes (Cotmore *et al.*, 2014; King *et al.*, 2011). The host of *Bombyx mori bidensovirus* (BmBDV), the only member of the *Bidnaviridae* family, is the silkworm *Bombyx mori* (Hu *et al.*, 2013). Collectively, animal-infecting viruses are associated with a multitude of pathologies including hepatitis, post-weaning multisystemic wasting syndrome, mink viral enteritis, and densonucleosis in *B. mori* (Allan *et al.*, 1998; Barreto *et al.*, 2014; Ellis *et al.*, 1998; Hu *et al.*, 2013; King *et al.*, 2011; Steinel *et al*., 2001). These four families have similar morphologies where the ssDNA genome is encapsidated in non-enveloped virions with T=1 icosahedral symmetry (Hu *et al.*, 2013; Itoh *et al.*, 2000; King *et al.*, 2011; Siegl *et al.*, 1971; Tischer *et al.*, 1982). The genomes of anelloviruses (2.0-3.9kb) and circoviruses (1.7-2.0kb) are circular, whereas parvoviruses (4.0-6.3kb) and bidnaviruses (6.5kb and 6kb) have linear genomes (Hu *et al.*, 2013; King *et al.*, 2011). BmBDV is the only virus with a bipartite genome that infects animals (Hu *et al.*, 2013). This virus was originally classified as a member of the *Densovirus* genus (*Parvoviridae* family) but due to its unique coding strategy and significantly larger genome it was assigned to a new family (Hu *et al.*, 2013; King *et al.*, 2011).

1.2.1.2 Plant-infecting ssDNA viruses

Geminiviridae and *Nanoviridae* are families of plant-infecting viruses with circular ssDNA genomes (King *et al.*, 2011). Geminivirus genomes are encapsidated in geminate particles and nanovirus virions are icosahedral (Harding *et al.*, 1991; King *et al.*, 2011; Zhang *et al.*, 2001). Members of the *Nanoviridae* family have multicomponent genomes that are thought to require at least six (*Babuvirus* genus) or eight (*Nanovirus* genus) 0.9-1.1kb ssDNA components in order to cause infection, with the majority of components encoding a single protein (Burns *et al.*, 1995; Grigoras *et al.*, 2009; King *et al.*, 2011). Each component of the multiple-component genome of the *Nanoviridae* family is encapsidated in separate virions that are transmitted by aphids (King *et al.*, 2011). The majority of geminiviruses have 2.5-3.0kb monopartite genomes (King *et al.*, 2011; Stanley, 1983), however, a number of members of the *Begomovirus* genus have bipartite genomes. These bipartite genomes

have two components, DNA-A and DNA-B, each 2.5-2.6kb in size, which are individually encapsidated. Transmission of geminiviruses is mediated by leafhoppers, plant hoppers, treehoppers and whiteflies (Hogenhout *et al.*, 2008; King *et al.*, 2011) and recent evidence suggests that aphids are able to transmit a new group of geminiviruses (Roumagnac *et al.*, 2015). Geminiviruses and nanoviruses infect both monocotyledonous and dicotyledonous plants and are the causative agents of disease in numerous crops including bananas, broad beans, chickpeas, cotton and maize (Grigoras *et al.*, 2009; Harding *et al.*, 1991; King *et al.*, 2011; Kumari *et al.*, 2008; Varma & Malathi, 2003). Satellite molecules are associated with some geminiviruses and nanoviruses (Briddon *et al.*, 2003; Horser *et al.*, 2001b; King *et al.*, 2011; Kumar *et al.*, 2013; Patil & Fauquet, 2010; Rohde *et al.*, 1990; Saunders *et al.*, 2000). Satellite molecules rely on the associated virus for replication, encapsidation and/or transmission. Alphasatellites are nanovirus-like molecules that encode a replication-associated protein (Rep) and hence replicons that can reduce disease symptoms, whereas betasatellites can be associated with enhancing disease symptoms (Briddon *et al.*, 2003; Kumar *et al.*, 2013; Saunders *et al.*, 2000; Stanley, 2004).

1.2.1.3 Prokaryote-infecting ssDNA viruses

Microviridae, *Inoviridae* and *Spiraviridae* are families of ssDNA viruses that infect prokaryotes. The prokaryote-infecting viruses have circular ssDNA genomes encapsidated within virions of differing morphology (King *et al.*, 2011; Mochizuki *et al.*, 2012). The *Microviridae* family is divided into the *Microvirus* genus (5.3-6.1kb genome), which infects enterobacteria, and the *Gokushovirinae* subfamily, which contains three genera (4.5-6.0kb genome) that infect obligate parasitic bacteria and *Spiroplasma* (King *et al.*, 2011). All members have non-enveloped virions with a T=1 icosahedral symmetry (Hsia *et al.*, 2000; King *et al.*, 2011). There are two genera within the *Inoviridae* family, *Inovirus* and *Plectrovirus*, whose circular ssDNA genomes (6.0-12.4kb) are encapsidated in filaments or short rod-like particles, respectively (King *et al.*, 2011; Yamada *et al.*, 2007). Inoviruses infect gram-positive and gram-negative bacteria and plectroviruses infect mycoplasmas (King *et al.*, 2011). Some inoviruses are associated with disease, mediating the horizontal transfer of virulence factors within a population of pathogenic bacteria (Bille *et al.*, 2005;

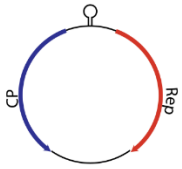
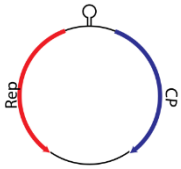
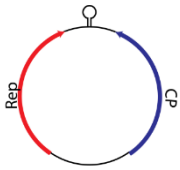
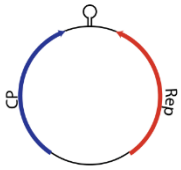
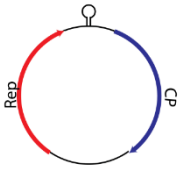
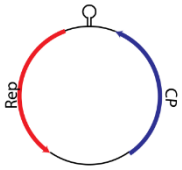
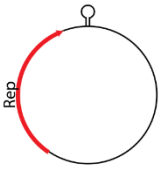
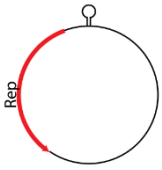
Campos *et al.*, 2010; King *et al.*, 2011). The only member of *Spiraviridae* is *Aeropyrum coil-shaped virus* (ACV), which infects an archaea, *Aeropyrum pernix* (Mochizuki *et al.*, 2012). ACV has a circular ssDNA genome encapsidated within a hollow, cylindrical helix that lacks an envelope (Mochizuki *et al.*, 2012). The genome of ACV is ~24.9kb, thus it is the largest of any ssDNA virus identified to date. ACV was assigned to a new family due to its unique genome and morphology.

1.2.2 Genome organisations of eukaryotic single-stranded DNA viruses that encode a replication-associated protein

SsDNA viruses are among the smallest eukaryotic pathogens (Rosario *et al.*, 2012b). SsDNA viruses have minimal genomes because they rely heavily on host machinery for their replication. Eukaryotic circular ssDNA viruses encode as few as two to six proteins (King *et al.*, 2011; Rosario *et al.*, 2012b). The circular genomes of prokaryotic viruses are larger and can encode up to seventeen or more genes (King *et al.*, 2011).

Well-characterised circular Rep-encoding single-stranded (CRESS) DNA viruses are circoviruses, geminiviruses and nanoviruses. The last couple of years have seen a significant increase in novel CRESS DNA viruses that cannot be classified within existing viral families. The Rep is essential as it initiates rolling circle replication of the circular genomes. Expression of both spliced and unspliced *rep* transcripts is required for regulation of gene expression, infectivity and replication in some members of the *Geminiviridae* family (Collin *et al.*, 1996; Dekker *et al.*, 1991; Liu *et al.*, 1998; Liu *et al.*, 1999b; Wright *et al.*, 1997). In bipartite begomoviruses and multipartite nanoviruses, the Rep is encoded by DNA-A and DNA-R components, respectively (Burns *et al.*, 1995; Horser *et al.*, 2001a; King *et al.*, 2011). Once expressed, the encoded Rep binds to iterons in common regions on the other genome components and mediates their replication (Timchenko *et al.*, 1999). The genome organisations differ among CRESS DNA viruses so a classification scheme was proposed (Rosario *et al.*, 2012b). CRESS DNA viruses are categorised according to orientation of the major open reading frames (ORFs), position of the stem-loop relative to the ORFs and orientation of the origin of replication with respect to the

Table 1.2: Classification of single-stranded DNA genomes where the nonanucleotide motif is located on the positive-sense strand. Adapted from Rosario *et al.* (2012b)

Genome organisation	Example illustration	ORF orientation	Nonanucleotide motif found on Rep-encoding strand	Orientation of ORFs relative to stem-loop
Type I		Ambisense	Yes	ORFs read away from the stem-loop
Type II		Ambisense	No	ORFs read away from the stem-loop
Type III		Ambisense	Yes	ORFs read towards the stem-loop
Type IV		Ambisense	No	ORFs read towards the stem-loop
Type V		Unisense	Yes	ORFs read clockwise
Type VI		Unisense	No	ORFs read counter-clockwise
Type VII		Single ORF	Yes	ORFs read clockwise
Type VIII		Single ORF	No	ORFs read counter-clockwise

Rep ORF (Table 1.2). Type I and type II genomes encode the Rep and capsid protein (CP) ORFs bidirectionally and feature a putative stem-loop structure located at the 5' end of the major ORFs. Type III and type IV viruses also have ambisense genomes, but the stem-loop structure is located at the 3' end of the major ORFs. Genomes with a unisense orientation are classified as type V or VI genomes and circular molecules containing a single ORF are classified as type VII or VIII genomes. The nonanucleotide sequence of type I, III, V and VII genomes is located on the Rep-encoding strand, but this is not true of type II, IV, VI and VIII genomes.

1.2.3 Replication of CRESS DNA viruses

CRESS DNA viruses replicate through the rolling circle replication (RCR) mechanism. This mechanism is shared by bacterial plasmids, indicating a possible evolutionary link between these entities (Khan, 1997; Koonin & Ilyina, 1992). The RCR mechanism occurs in three stages: initiation, elongation and termination (Gilbert & Dressler, 1968; Gutierrez, 1999; Martin *et al.*, 2011a; Rosario *et al.*, 2012b). The RCR mechanism was determined using experimental evidence from geminivirus studies, however, structural analysis of Reps suggest that nanoviruses and circoviruses replicate through a similar mechanism (Cheung, 2012; Hafner *et al.*, 1997; Rosario *et al.*, 2012b; Timchenko *et al.*, 1999; Vega-Rocha *et al.*, 2007a; Vega-Rocha *et al.*, 2007b).

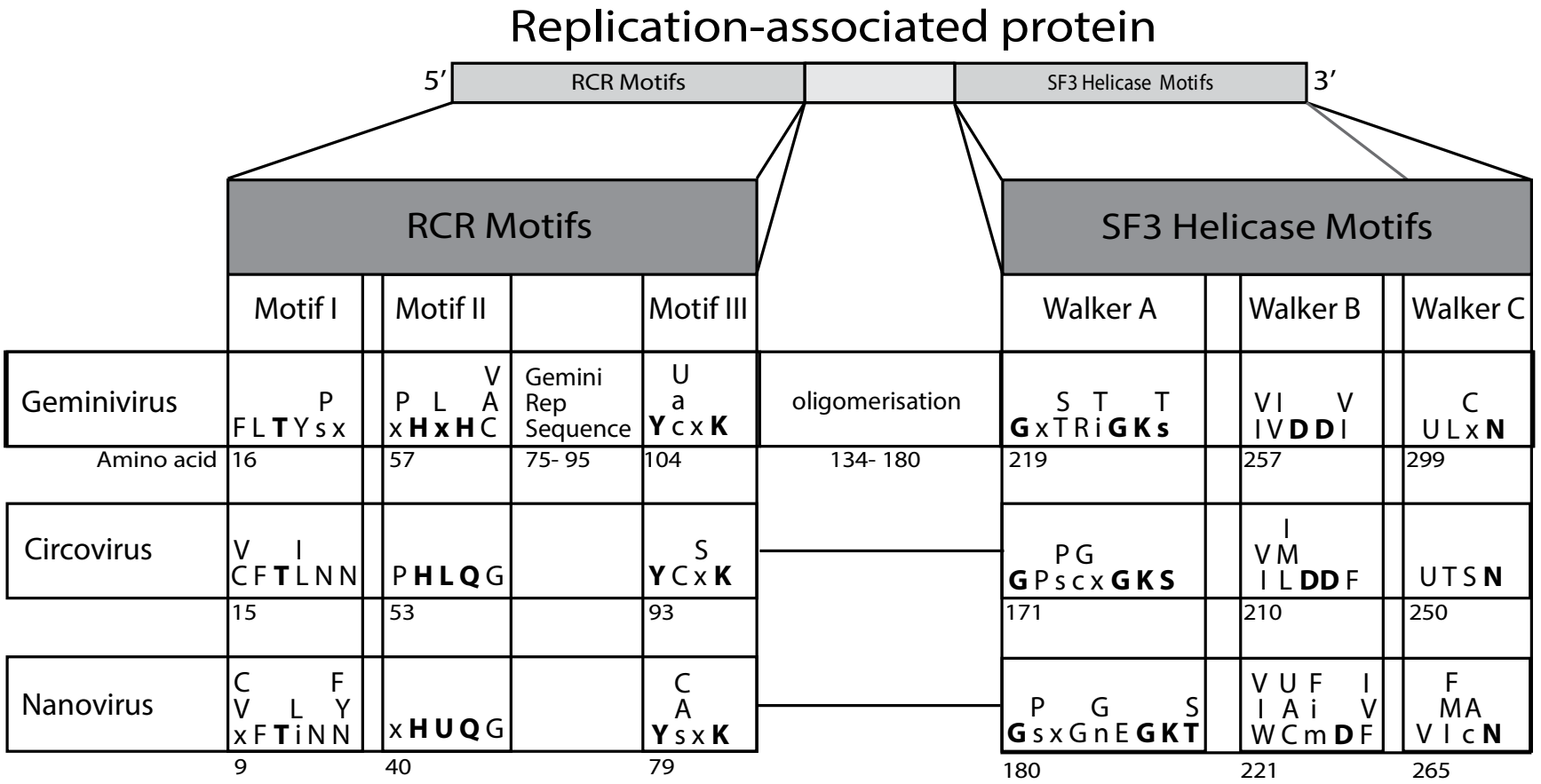
Viral DNA is localised to the nucleus by the CP following infection (Kunik *et al.*, 1998; Liu *et al.*, 1999a). A priming event is initiated and the circular ssDNA genome is converted into a covalently closed double-stranded DNA (dsDNA) intermediate by host polymerases (Gutierrez, 1999; Martin *et al.*, 2011a). The dsDNA interacts with histone proteins and is coiled into mini-chromosomes (Abouzid *et al.*, 1988; Paprotka *et al.*, 2015; Pilartz & Jeske, 1992). The Rep cleaves the dsDNA between positions 7 and 8 of the nonanucleotide sequence (NANTNTTAN) located at a putative stem-loop structure at the origin of replication (Cheung, 2004; Hafner *et al.*, 1997; Heyraud-Nitschke *et al.*, 1995; Laufs *et al.*, 1995b; Steinfeldt *et al.*, 2006; Timchenko *et al.*, 1999). This produces a circular replicative form that is elongated at the exposed 3' end by host machinery and Rep helicase activity in a continuous cyclical manner.

This gradually displaces the template strands (Gutierrez, 1999; Jeske *et al.*, 2001; Martin *et al.*, 2011a; Saunders *et al.*, 1991). After one or more cycles, the displaced strands are ligated, forming circular monomeric or multimeric ssDNA virions (Hafner *et al.*, 1997; Jeske *et al.*, 2001; Laufs *et al.*, 1995b; Martin *et al.*, 2011a; Steinfeldt *et al.*, 2006; Timchenko *et al.*, 1999). The circular DNA molecules can be encapsidated into virions if monomeric, or enter subsequent rounds of replication (Jeske *et al.*, 2001; Martin *et al.*, 2011a).

1.2.3.1 N-terminal domain of the replication-associated protein

The Rep N-terminus of CRESS DNA viruses contain motifs that are important in initiating RCR (Figure 1.1). The motifs are well conserved across ssDNA viruses, phage and plasmids that replicate using the RCR mechanism (Koonin & Ilyina, 1992; Kraberger *et al.*, 2015a; Rosario *et al.*, 2012b; Vega-Rocha *et al.*, 2007a; Vega-Rocha *et al.*, 2007b). The Rep N-terminus also contains specificity binding determinants (SPDs) that interact to form a beta-sheet, permitting high affinity binding between the Rep and dsDNA intermediate (Argüello-Astorga & Ruiz-Medrano, 2001; Londoño *et al.*, 2010; Mauricio-Castillo *et al.*, 2014; Orozco & Hanley-Bowdoin, 1998). The conserved sequence of RCR motif I is Fu(t/u)(l/y)(t/p), where 'u' denotes a bulky hydrophobic residue (Rosario *et al.*, 2012b). Due to its proximity with SPDs, RCR motif I is thought to be involved in the recognition of iterative sequences associated with the origin of replication (Argüello-Astorga & Ruiz-Medrano, 2001; Rosario *et al.*, 2012b). The exact function of RCR motif I is undetermined. The conserved amino acid sequence of RCR motif II is (p/u)/HuH in geminiviruses and (p/u)HuQ in circoviruses and nanoviruses (Rosario *et al.*, 2012b). Conserved histadine residues coordinate divalent metal ions, Mg^{2+} or Mn^{2+}, which are important cofactors for endonuclease activity at the origin of replication (Ilyina & Koonin, 1992; Laufs *et al.*, 1995b). In circoviruses and nanoviruses, it is suggested that acidic side chains with a similar spatial positioning to the second conserved histadine residue in RCR motif II of geminiviruses coordinate the divalent metal ions (Vega-Rocha *et al.*, 2007a; Vega-Rocha *et al.*, 2007b). An invariant tyrosine residue in RCR motif III (YxxK, where 'x' denotes any amino acid) catalyses dsDNA cleavage and subsequent covalent

Figure 1.1: Conserved RCR and SF3 helicase motifs in the replication-associated protein of circoviruses, geminiviruses and nanoviruses [adapted from Rosario *et al.* (2012b)]. Conserved sequences are annotated as follows: invariant residues are in bold, uppercase residues occur at a higher frequency than residues denoted by lowercase letters, 'x' denotes any residue, and 'U' denotes any bulky hydrophobic residue (I, L, V, M, F, Y or W). The position of conserved motifs in representative species of *Circoviridae* (*Porcine circovirus* 1; NC_001792), *Geminiviridae* (*Tomato golden mosaic virus*; NC_001507) and *Nanoviridae* (*Faba bean necrotic yellows virus*; NC_003560) is indicated.



attachment of Rep to the 5' end of the cleaved product (Laufs *et al.*, 1995a; Laufs *et al.*, 1995b; Orozco & Hanley-Bowdoin, 1998; Rosario *et al.*, 2012b; Steinfeldt *et al.*, 2007; Timchenko *et al.*, 1999). The presence of a single tyrosine residue in RCR motif III classifies geminivirus, circovirus and nanovirus Reps as superfamily II Reps (Ilyina & Koonin, 1992). The conserved lysine residue in RCR motif III is proposed to mediate binding and positioning during catalysis (Vega-Rocha *et al.*, 2007a; Vega-Rocha *et al.*, 2007b). A fourth conserved motif, the geminivirus Rep sequence (GRS), is only found in geminiviruses and a group of CRESS DNA viruses called gemycircularviruses (Dayaram *et al.*, 2012; Nash *et al.*, 2011). It enables appropriate spatial arrangements of RCR motifs II and III (Nash *et al.*, 2011). Site directed mutagenesis of the GRS domain in *Tomato golden mosaic virus* yielded non-infectious clones, demonstrating that the GRS is essential for geminivirus, and presumably gemycircularvirus, replication (Dayaram *et al.*, 2012; Nash *et al.*, 2011).

1.2.3.2 Helicase domain of the replication-associated protein

Rep is a multifunctional protein, possessing both endonuclease and helicase activities. Rep helicase activity is mediated by conserved motifs in a C-terminal NTP-binding domain, named Walker A, Walker B, motif B' and motif C (Figure 1.1) (Choudhury *et al.*, 2006; Clérot & Bernardi, 2006; Gorbalenya *et al.*, 1990; Koonin, 1993). The conserved motifs are found within a 120 nucleotide (nt) long sequence, classifying encoded Reps as superfamily 3 (SF3) helicases (Gorbalenya *et al.*, 1990; Koonin, 1993). During synthesis of progeny strands, Rep helicase activity unwinds the dsDNA intermediate in the 3' to 5' direction using nucleotide triphosphates (Choudhury *et al.*, 2006; Clérot & Bernardi, 2006). The conserved Walker A motif [GxxxxGK(S/T)] forms part of a P-loop structure in the NTP-binding domain that facilitates ATP recognition and binding with a conserved lysine residue (Choudhury *et al.*, 2006; Clérot & Bernardi, 2006; Desbiez *et al.*, 1995; George *et al.*, 2014; Rosario *et al.*, 2012b; Timchenko *et al.*, 1999). Site directed mutagenesis demonstrates that the Walker B motif [hhxh(D/E)(D/E), where 'h' represents a hydrophobic residue] contributes to ATP binding and is essential in ATP hydrolysis, while motif C [h(T/S/x)(T/S/x)N] interacts with the gamma phosphate of ATP and the nucleophilic water molecule via a conserved asparagine residue (Choudhury *et al.*, 2006; George *et*

al., 2014). The fourth motif, motif B' [(K/R)x_{3-4}Gx_{7-8}K], is specific to SF3 helicases and is found between Walker B and motif C (George *et al.*, 2014; Koonin, 1993; Yoon-Robarts *et al.*, 2004). It is suggested that this motif promotes allosteric changes in the Rep to mediate interactions with the dsDNA intermediate for ATPase and helicase activity (George *et al.*, 2014; Yoon-Robarts *et al.*, 2004).

1.2.4 Evolution of single-stranded DNA viruses

1.2.4.1 Genetic drift

The rate of ssDNA virus evolution is comparable to RNA viruses. Specifically, monopartite geminiviruses evolve at a rate on the order of $2.0x10^{-4}$ - $3.5x10^{-4}$ substitutions per site per year and bipartite geminiviruses evolve at mean rates of $1.33x10^{-4}$ and $1.6x10^{-3}$ substitutions per site per year for DNA-A and DNA-B components, respectively (Duffy & Holmes, 2008; 2009; Harkins *et al.*, 2009). A similar mutation rate of $1.78x10^{-3}$ substitutions per site per year was also observed in a nanovirus, *Faba bean necrotic stunt virus* (Grigoras *et al.*, 2010). Estimated rates of $0.520x10^{-3}$ - $1.226x10^{-3}$ substitutions per site per year were recorded for the CP of *Porcine circovirus* 2 (Nguyen *et al.*, 2012), while the CP and Rep of *Porcine parvovirus* have estimated rates of $3.02x10^{-4}$ - $5.39x10^{-5}$ substitutions per site per year, respectively (Streck *et al.*, 2011). Similarly, mean evolutionary rates of $5.29x\ 10^{-4}$ - $5.51x10^{-4}$ substitutions per site per year were recorded for *Torque teno sus viruses* 1 and 2 (*Anelloviridae*) (Cadar *et al.*, 2013).

High mutation rates are unexpected as ssDNA viruses are replicated by host polymerases with presumed high fidelity and proofreading capabilities. Biased substitution patterns indicate possible mechanisms by which this high mutation rate occurs. High rates of C→T and G→A substitutions were reported by Duffy & Holmes (2008), suggesting that prolonged exposure of unpaired bases results in enzymatic or spontaneous deamination events on the ssDNA genome. Additionally, G→T transversions are overrepresented in *Maize streak virus* populations (van der Walt *et al.*, 2008). This is indicative of increased guanine oxidation or decreased repair of oxidative damage. The impairment of other repair mechanisms may also promote

increased mutation rates. Evidence suggests that the DNA of geminiviruses is unmethylated, therefore mismatch repair mechanisms may not function correctly during replication (Brough *et al.*, 1992; Ge *et al.*, 2007; Inamdar *et al.*, 1992; Roossinck, 1997). It is also suggested that base-excision repair may not operate because dsDNA intermediates are transient during RCR (Duffy & Holmes, 2008).

1.2.4.2 Recombination and reassortment

SsDNA viruses also explore sequence space through recombination and reassortment. Evidence of recombination events in anelloviruses, circoviruses, geminiviruses and associated satellite molecules, microviruses, nanoviruses and parvoviruses (Amin *et al.*, 2006; Cadar *et al.*, 2013; Cai *et al.*, 2012; Cheung, 2009; Grigoras *et al.*, 2014; Huang *et al.*, 2013; Idris & Brown, 2004; Julian *et al.*, 2013; Lefeuvre *et al.*, 2009; Leppik *et al.*, 2007; Martin *et al.*, 2011b; Mochizuki *et al.*, 2008; Rokyta *et al.*, 2006; Savory & Ramakrishnan, 2014; Silva *et al.*, 2014; Stainton *et al.*, 2012; Tyumentsev *et al.*, 2014; Varsani *et al.*, 2008), and reassortment events in nanoviruses have been found (Grigoras *et al.*, 2014; Savory & Ramakrishnan, 2014; Stainton *et al.*, 2012; Stainton *et al.*, 2015). Recombination is the exchange of homologous or non-homologous sequences between genomes. This requires replication of more than one genome within the same host cell, therefore recombination rates are maximised by shared geographical distributions, epidemiology and tissue tropism (Martin *et al.*, 2011a).

The exact mechanism underlying recombination in ssDNA viruses is not clear. Homologous recombinants may form through a copy-choice mechanism, whereby replication complexes disassociate and replication resumes at a different location (Martin *et al.*, 2011a; Viguera *et al.*, 2001). Template switching is likely promoted by secondary structures that stall the replication complexes, such as the stem-loop at the origin of replication which is a known recombination hotspot (Lefeuvre *et al.*, 2009). Replication complexes may also dissociate due to clashes with transcription machinery (Helmrich *et al.*, 2013; Martin *et al.*, 2011a). Alternatively, recombinants may be formed by double-stranded break repair mechanisms (Cromie *et al.*, 2001; Xu & Price, 2011). Resection causes the formation of single-stranded 3’ overhangs at the

double-stranded break site. Recombinants are formed by hybridisation between the 3' overhang and homologous sequences, priming synthesis of a complementary strand. Covalently closed circular dsDNA intermediates can be replicated multiple times by host polymerases, producing heterogenous-length high molecular weight double-stranded DNA (hDNA) that is likely released as unit-length genomes (Alberter *et al.*, 2005; Jeske *et al.*, 2001; Martin *et al.*, 2011a; Stenger *et al.*, 1991). Break resolution by non-homologous end joining involves degradation of the 3' overhangs by nuclease activity and subsequent ligation to form non-homologous recombinants (Lieber *et al.*, 2003).

Reassortment, or pseudo-replication, involves the exchange of whole components between isolates. Reassortment has been documented in the multipartite genomes of nanoviruses, bipartite begomoviruses (*Geminiviridae*) and associated satellite molecules (Brown *et al.*, 2002; Grigoras *et al.*, 2014; Idris & Brown, 2004; Malik *et al.*, 2011; Silva *et al.*, 2014; Stainton *et al.*, 2012; Stainton *et al.*, 2015; Unseld *et al.*, 2000). Studies suggest that these genome reassortments can yield infectious clones with distinct biological properties relative to the parental viruses (Brown *et al.*, 2002; Chakraborty *et al.*, 2008). This process is restricted, as fully functional viruses must inherit components that will cooperate to ensure successful viral propagation (Hill *et al.*, 1998; Sung & Coutts, 1995). Inter-component homologous recombination frequently occurs between common regions to maintain *trans*-replicational control of reassorted components (Hu *et al.*, 2007; Hughes, 2004; Martin *et al.*, 2011a; Savory & Ramakrishnan, 2014; Stainton *et al.*, 2012). The common region of a newly introduced component can be replaced by that of the DNA-A / DNA-R component during the capture process, however a period of adaptive evolution may be required to optimise functionality (Hou & Gilbertson, 1996; Hu *et al.*, 2007; Hughes, 2004; Jovel *et al.*, 2007; Jovel *et al.*, 2004; Saunders *et al.*, 2002). The evolution of multipartite ssDNA genomes is thought to involve recombination between the DNA-A component of geminiviruses and associated satellite molecules (Briddon *et al.*, 2010; Martin *et al.*, 2011a).

Recombination events occur between viral and host genomes, as well as between co-replicating viruses (Frischmuth & Stanley, 1998; Martin *et al.*, 2011a; Saunders &

Stanley, 1999; van der Walt *et al.*, 2009). While recombination provides advantages to participating viral genomes, including rescue of defective molecules and rapid exploration of adaptive sequences, some recombinant genomes are less fit (Davino *et al.*, 2009; Jeske *et al.*, 2001; Martin *et al.*, 2011a). Reduced fitness may be caused by disruptions to interactions within the ssDNA genome, secondary structure formation, and protein tertiary and quaternary structures.

1.3 Approaches for the discovery of novel single-stranded DNA viruses

1.3.1 Phi29 DNA polymerase

Phi29 DNA polymerase has played a significant role in the discovery of novel circular DNA viruses (Johne *et al.*, 2009). The genetic material of ssDNA viruses can be amplified by sequence-independent rolling circle amplification (RCA) prior to next-generation sequencing (NGS). RCA uses bacteriophage Phi29 DNA polymerase, a highly processive enzyme with 3' - 5' exonuclease and proofreading activities (Blanco *et al.*, 1989; Garmendia *et al.*, 1992). Concentrations exceeding 0.1ρg of DNA are sufficiently amplified by Phi29 DNA polymerase, with a reported error rate of $3x10^{-6}$ - $5x10^{-6}$ (Nelson *et al.*, 2002). Phi29 DNA polymerase possesses strand-displacement properties, permitting exponential non-specific amplification of circular DNA templates with random hexamers (Blanco *et al.*, 1989; Nelson *et al.*, 2002). Circular templates are preferentially amplified, however, linear DNA is also efficiently amplified by the polymerase (Nelson *et al.*, 2002).

1.3.2 Next-generation sequencing platforms

Over the past decade, a number of NGS platforms have been developed in response to limitations of Sanger sequencing (Metzker, 2010). NGS platforms enable cost-effective acquisition of a large amount of data in a shorter timeframe than Sanger sequencing. NGS platforms have been used successfully in fields including genetics, medicine, metagenomics and evolution (Metzker, 2010; Rothberg & Leamon, 2008). NGS platforms generally involve library preparation, sequencing and imaging

followed by *de novo* or scaffold based assembly of shorter contigs into the complete DNA sequence (Metzker, 2010). The library is prepared by shearing DNA sequences to generate random fragments (Shendure & Ji, 2008). Alternatively, a library of mate-paired reads can be used. Adapter sequences are ligated to each end of the DNA to immobilise fragments onto a surface for sequencing. Universal primers are used during amplification steps, permitting sequencing of target DNA without prior knowledge of the sequence. The processes and chemistry used for sequencing and imaging differ between NGS platforms.

The following is a brief discussion of selected second- and third-generation platforms: Roche 454 pyrosequencing, Illumina/Solexa sequencing, SOLiD sequencing, Ion Torrent sequencing, single-molecule real-time sequencing with PacBio, and nanopore sequencing platforms.

1.3.2.1 Roche 454 pyrosequencing

Roche 454 pyrosequencing was the first NGS platform available commercially (Margulies *et al.*, 2005). A library of random adapter-flanked DNA fragments is used. Individual fragments are bound to 28μm beads that are captured inside an oil emulsion (Margulies *et al.*, 2005; Shendure & Ji, 2008). Emulsion PCR inside each oil droplet produces clonal populations of amplified fragments. The oil droplet is broken and unbound complementary strands are washed away by a denaturant. Amplicon-bearing beads are first incubated with *Bacillus stearothermophilus* polymerase, followed by incubation with a single-stranded binding protein (Shendure & Ji, 2008). The incubated beads are then deposited into individual picolitre-scale wells on a fibre-optic slide with beads containing luciferase and ATP sulfurylase (Margulies *et al.*, 2005; Shendure & Ji, 2008). A modified pyrosequencing protocol is implemented on individual fibre-optic slides containing approximately 1.6 million wells. Unlabelled deoxynucleotides (dNTPs) are added one at a time, followed by an apyrase wash after each cycle to ensure that the nucleotide is sufficiently removed before addition of the next dNTP species. A high reagent volume is therefore required. Incorporation of a nucleotide into the complementary strand releases a pyrophosphate molecule that is used to generate light by ATP sulfurylase and

luciferase activities. Photon release is detected by the charge-coupled device attached to the bottom of the fibre-optic slide. Signal intensity relative to a baseline level is used to determine if a nucleotide(s) was incorporated into the complementary strand. The pattern of photon release over time is used to elucidate the sequence of a DNA fragment at a given coordinate in the microarray. This approach is asynchronous, as the rate of base addition to the complementary strand will differ between fragments depending on their sequence. Interpretation is complicated by homopolymers, as consecutive incorporations of a dNTP species in a single cycle is not prevented (Shendure & Ji, 2008). Therefore, the insertion-deletion error rate in this approach is higher than other sequencing technologies. High-quality reads with distinct differences between baseline levels and an incorporation event are selected to determine the sequence of the DNA fragment (Margulies *et al.*, 2005). The high-quality reads are then aligned to determine the sequence of the target DNA molecule. The read length supported by Roche 454 pyrosequencing is higher than other established NGS platforms (Shendure & Ji, 2008).

1.3.2.2 Illumina/Solexa sequencing

Illumina/Solexa sequencing utilises a sequencing-by-synthesis approach (Metzker, 2010; Shendure & Ji, 2008). DNA fragments that are approximately 200-300nt in size are selected from a randomly generated library. Adapter sequences are ligated to each end of the fragments and hybridise with forward and reverse primers attached on the inside surface of flow cell channels. Clonal populations of the immobilised fragments are generated within independent lanes by bridge amplification. A single flow-cell contains eight lanes, facilitating simultaneous sequencing of up to eight independent DNA libraries. Following amplification, the dsDNA is denatured by formamide and the reverse strands and PCR components are washed away. DNA polymerase, sequencing primers and nucleotides with differently coloured fluorophores are added to the flow cell for synthesis of complementary strands. The labelled nucleotide acts as a 3'-blocked reversible terminator to ensure that nucleotides are incorporated into the complementary strand in separate events. This approach therefore overcomes the homopolymer issues of Roche 454 pyrosequencing (Shendure & Ji, 2008). A laser excites the bound nucleotide's fluorophore and the fluorescent pulse is recorded by

the optical system. The fluorophore is chemically removed, priming the strand for further incorporation events. Base-calling is informed by emission wavelength and pulse intensity and a quality checking pipeline ensures high raw-read accuracy. Once base-calling has identified the sequences of DNA fragments, the short reads are aligned to determine the sequence of the complete DNA molecule. This is a high throughput method, generating 100-200 million copies of the original template that are sequenced in parallel (Metzker, 2010). However, data acquisition takes more time than other NGS platforms and shorter reads are generated (Metzker, 2010; Quail *et al.*, 2012). The predominant errors in Illumina sequencing are substitution errors. As sequencing proceeds, an increased substitution error rate occurs due to incomplete removal of unincorporated fluorescent nucleotides between events, resulting in higher background noise.

1.3.2.3 SOLiD system

Support oligonucleotide ligation detection (SOLiD), developed by Applied Biosystems, is a sequencing-by-ligation NGS platform (Metzker, 2010; Shendure & Ji, 2008). This approach involves emulsion PCR where a library of fragmented DNA is attached to microreactors containing template DNA, PCR reaction components, 1μm paramagnetic beads and primers. This creates clonal populations of DNA templates attached to separate beads. Simultaneous PCR reactions are followed by denaturation and selection of beads with extended DNA templates. Selected beads undergo a 3' modification and are covalently attached to a flow cell. Approximately 100 million reads are generated per flow cell (Rothberg & Leamon, 2008), although these reads are very short and the data acquisition process takes longer than other NGS platforms (Metzker, 2010; Schadt *et al.*, 2010). Universal primers anneal to adapter sequences on each bead and di-base octamer probes are ligated to the primers using DNA ligase. The probes identify a two-base combination, thus there are sixteen possible dinucleotide sequences encoded by four dyes of different colours. Distinct probes corresponding to four of the dinucleotide sequences are added at a time, where the complementary di-base probe hybridises to the amplified template and is ligated to the primer. Upon ligation, the fluorescent signal is emitted. The fluorophore is then cleaved using silver ions to allow attachment of subsequent probes. A cycle of probe

hybridisation, ligation to the template DNA, imaging and cleavage occurs, with every fifth base sequenced. The extended primer is removed and a new universal primer is added complementary to the n-1 position. Primer reset allows sequencing of the next frame using the cycle described. SOLiD sequencing achieves high quality sequencing as the identity of each base is confirmed in multiple cycles. The short reads are aligned to determine the DNA sequence of the unfragmented DNA molecule.

1.3.2.4 Ion Torrent sequencing

Ion Torrent sequencing is a sequencing-by-synthesis approach that uses an efficient complementary metal-oxide-semiconductor (CMOS) chip (Merriman *et al.*, 2012). Both sequencing and data collection is performed on the CMOS chip which contains millions of individual transistor-based sensors. This allows extensive sequencing of DNA molecules in parallel. DNA molecules from a library of fragments are bound onto individual beads that are deposited inside microwells. A transistor-based sensor is also located at the bottom of each microwell. Clonal populations of fragmented sample DNA are generated by emulsion PCR, in a manner similar to Roche 454 pyrosequencing (see section 1.3.2.1). dNTPs are sequentially added to the biosensor array followed by a wash step to ensure each dNTP is completely removed before the next species is added. Incorporation of the dNTP into the complementary strand causes a release of pyrophosphate molecules and hydrogen ions. The latter is detected by the pH-sensitive field effect transistor (pHFET) device. An incorporation event is recorded by the pHFET device as a change in current through the transistor. The change in current is proportional to the number of incorporation events per round, allowing homopolymeric templates to be sequenced more accurately than Roche 454 pyrosequencing, however the Ion Torrent sequencing platform fails to sequence extended homopolymeric templates (Quail *et al.*, 2012). A base-calling algorithm is used to convert the series of individual signals into the sequence of fragmented DNA (Merriman *et al.*, 2012). These short reads are then assembled into the complete target DNA sequence. Unlike Illumina/Solexa sequencing, this platform does not involve incorporation of labelled nucleotides into an elongating strand. Because unmodified nucleotides are used, this platform has reduced reagent costs and can support longer reads than some other NGS platforms (Merriman *et al.*, 2012; Quail *et al.*, 2012).

1.3.2.5 Single-molecule real-time sequencing

Single-molecule real-time (SMRT) sequencing, developed by Pacific Biosystems, is a third-generation NGS platform (Eid *et al.*, 2009; Korlach *et al.*, 2010; Schadt *et al.*, 2010). Individual Phi29 DNA polymerases are immobilised onto zero-mode waveguide (ZMW) nanostructures using biotin/streptavidin interactions. Each ZMW nanostructure acts as a nanophotonic visualisation chamber for sequencing of a single DNA molecule (Levene *et al.*, 2003). Arrays consist of approximately 3000 ZMW nanostructures, so this approach has a moderate throughput relative to other technologies (Metzker, 2010; Schadt *et al.*, 2010). A laser illuminates the bottom 30nm of the ZMW so that the SMRT DNA sequencing optical system can detect fluorescence. Base-calling is informed by fluorophores of different emission wavelengths that are attached to the terminal phosphate of nucleotides. The different phospholinked nucleotides are added to the ZMV nanostructure individually where they emit a fluorescence signal when stimulated by the laser. When the cognate nucleotide is added, it binds to the active site of the polymerase and the fluorophore is cleaved. The length of the fluorescent pulse is approximately three orders of magnitude longer when the nucleotide binds to the active site of the polymerase than simple diffusion. This change in time is detected as a higher intensity signal by the optical system. Synthesis of the complementary strand is recorded as single incorporation events, allowing the DNA sequence to be determined. As SMRT sequencing analyses single molecules, it allows much higher resolution sequencing and longer reads than other approaches, although data acquisition takes more time (Metzker, 2010; Quail *et al.*, 2012; Schadt *et al.*, 2010). Notably, this platform also displays a significant bias against AT-rich genomes and higher error rates than some other NGS platforms.

1.3.2.6 Nanopore sequencing technologies

Nanopore sequencing is an emerging NGS platform (Schadt *et al.*, 2010). This approach does not require modifications to the DNA molecule, amplification or involved chemical reactions (Branton *et al.*, 2008; Schadt *et al.*, 2010; Venkatesan & Bashir, 2011). Nanopores are set in graphene or insulating membranes that separate

electrolyte-filled chambers (Merchant *et al.*, 2010; Schneider *et al.*, 2010). A voltage is applied to drive translocation of unmodified DNA molecules through the nanopore. The diameter of biological or synthetic nanopores is comparable to DNA to ensure that molecules pass through the nanopore individually. Translocation of the DNA molecule disrupts the ionic current. The difference in current is dependent on the dNTP in the nanopore. The DNA sequence can therefore be determined by recording changes in ionic current over time. A sensor array chip is used consisting of microscaffolds each with a membrane, nanopore and electrode. This permits high throughput sequencing of DNA molecules. This approach supports longer read lengths than other sequencing technologies and requires low reagent volumes. Similar nanopore sequencing technologies have been developed that use a transistor to reduce the translocation speed, reducing base-calling errors (Schadt *et al.*, 2010).

1.4 Aims and rationale of this study

Viral metagenomic approaches have identified novel CRESS DNA viruses from various environmental samples, including air (Roux *et al.*, 2013; Whon *et al.*, 2012), wastewater / sewage (Cantalupo *et al.*, 2011; Kraberger *et al.*, 2015a; Ng *et al.*, 2012; Phan *et al.*, 2015; Rosario *et al.*, 2009b; Roux *et al.*, 2013), soil (Kim *et al.*, 2008; Reavy *et al.*, 2015), water ecosystems (Breitbart *et al.*, 2015; Dayaram *et al.*, 2014; Dayaram *et al.*, 2016; Dayaram *et al.*, 2015a; Fahsbender *et al.*, 2015; Hewson *et al.*, 2013a; Hewson *et al.*, 2013b; Kim *et al.*, 2015; Labonté & Suttle, 2013; López-Bueno *et al.*, 2009; Ng *et al.*, 2013; Rosario *et al.*, 2009b; Rosario *et al.*, 2015a; Roux *et al.*, 2012; Smith *et al.*, 2013; Yoshida *et al.*, 2013; Zawar-Reza *et al.*, 2014), plant material (Basso *et al.*, 2015; Dayaram *et al.*, 2012; Du *et al.*, 2014; Kraberger *et al.*, 2015b; Male *et al.*, 2015; Marzano & Domier, 2015) and invertebrates (Dayaram *et al.*, 2014; Dayaram *et al.*, 2013c; Dayaram *et al.*, 2015b; Garigliany *et al.*, 2015; Ng *et al.*, 2011b; Padilla-Rodriguez *et al.*, 2013; Pham *et al.*, 2013; Rosario *et al.*, 2012a; Rosario *et al.*, 2015a; Rosario *et al.*, 2015b). Notably, CRESS DNA viruses were recovered from animal or human faeces in numerous studies (Blinkova *et al.*, 2010; Castrignano *et al.*, 2013; Cheung *et al.*, 2014a; Cheung *et al.*, 2014b; Cheung *et al.*, 2013; Cheung *et al.*, 2015; Conceicao-Neto *et al.*, 2015; Garigliany *et al.*, 2014; Ge *et*

al., 2011; Ge *et al.*, 2012; Hansen *et al.*, 2015; Kim *et al.*, 2014; Kim *et al.*, 2012; Li *et al.*, 2010a; Li *et al*., 2013; Li *et al.*, 2011; Li *et al.*, 2010b; Li *et al.*, 2015; Lima *et al*., 2015; Male *et al.*, 2016; Ng *et al.*, 2014; Ng *et al.*, 2012; Ng *et al*., 2015; Phan *et al*., 2016; Phan *et al.*, 2011; Phan *et al.*, 2015; Phan *et al.*, 2013; Reuter *et al.*, 2014; Sachsenröder *et al.*, 2014; Sachsenröder *et al.*, 2012; Sasaki *et al.*, 2015; Sato *et al*., 2015; Shan *et al.*, 2011; Sikorski *et al.*, 2013a; Sikorski *et al.*, 2013b; Sikorski *et al.*, 2013d; Smits *et al*., 2014; Tan *et al*., 2013 ; van den Brand *et al.*, 2012; Victoria *et al.*, 2009; Woo *et al.*, 2014; Wu *et al.*, 2015; Zhang *et al.*, 2014). The recovered genomes represent members of the human or animal virome that are shed into faecal matter and the viral communities present in faeces, including associated prokaryotes, fungi, plant material and invertebrates. Sampling viruses in an ecosystem using faeces is of particular interest as it has potential as a surveillance tool for the management of disease outbreaks. Faecal sampling is an ideal system as it is non-invasive and cost-effective. The long life span of animals permits sampling of the viral reservoir for a longer time period with a broader scope than invertebrates where the viral reservoir concept has also been applied (Ng *et al.*, 2011b).

Many of the novel CRESS DNA viruses identified cannot be classified into viral taxonomic classifications. CRESS DNA viruses have been recovered from a range of sample types collected in New Zealand (Table 1.3). These viruses were isolated from animal faeces (Sikorski *et al.*, 2013a; Sikorski *et al.*, 2013b; Sikorski *et al.*, 2013d), bird blood and feathers (Jackson *et al.*, 2014a; Jackson *et al.*, 2015; Massaro *et al.*, 2012; Ortiz-Catedral *et al.*, 2010), sediment (Dayaram *et al.*, 2015a; Dayaram *et al*., 2016; Kraberger *et al.*, 2013), dragonflies and dragonfly larvae (Dayaram *et al.*, 2014; Dayaram *et al*., 2016; Dayaram *et al.*, 2013c), grass (Kraberger *et al.*, 2015b), molluscs (Dayaram *et al*., 2016; Dayaram *et al.*, 2015a; Dayaram *et al.*, 2013a; b), nesting material (Sikorski *et al.*, 2013c), lake water (Dayaram *et al*., 2016) and a sewage oxidation pond sample (Kraberger *et al.*, 2015a). With the exception of *Beak and feather disease virus* isolates (Jackson *et al.*, 2014a; Jackson *et al.*, 2015; Massaro *et al.*, 2012; Ortiz-Catedral *et al.*, 2010) and *Starling circovirus* (Dayaram *et al.*, 2013a), the remaining recovered CRESS DNA viruses are divergent and hence are not classified into any formal taxa. Some CRESS DNA viruses are clustered into proposed groupings due to sequence similarities and shared genome organisations,

namely the cycloviruses, smacoviruses and gemycircularviruses (Li *et al.*, 2011; Ng *et al.*, 2015; Rosario *et al.*, 2012a).

Additionally, various circular molecules have been discovered encoding a single ORF (Kraberger *et al.*, 2015a). The subgenomic DNA molecules encoded a Rep (n=8) or a CP (n=3) and may represent single components of multipartite CRESS DNA viruses or defective viral genomes that are an artefact of CRESS DNA virus replication.

Overall, the true extent of viral diversity in New Zealand, particularly of CRESS DNA viruses, is poorly understood. New Zealand CRESS DNA viruses have been identified largely due to the efforts of a single research group. Given the wide diversity in location and isolation source of recovered isolates, CRESS DNA viruses are likely to be highly prevalent in New Zealand. The objective of this study was to expand the limited knowledge of CRESS DNA virus diversity in New Zealand using faecal matter collected from various animal species.

1.4.1 Specific aims

The primary goal of this dissertation was to expand our knowledge of CRESS DNA viruses circulating in New Zealand using viral metagenomic approaches. Specifically, this study sought to:

1) Identify CRESS DNA viruses associated with faecal matter of domestic and wild animals in New Zealand
2) Determine, if any, the association of CRESS DNA virus species with faecal sources

The association of CRESS DNA viruses with animal faecal matter collected in New Zealand has not been extensively studied (Sikorski *et al.*, 2013a; Sikorski *et al.*, 2013b; Sikorski *et al.*, 2013d). Initial studies have demonstrated that diverse CRESS DNA viruses can be identified from the faeces of various animal species across the South Island and Chatham Islands of New Zealand. Namely, metagenomic analyses using animal faecal matter have identified fourteen gemycircularvirus isolates from

the faeces of multiple animal species, an unclassified CRESS DNA virus recovered from New Zealand fur seal faeces, and a smacovirus isolate from pig faeces. Furthermore, thirteen gemycircularvirus isolates, 37 unclassified CRESS DNA viruses and eleven circular molecules were identified in a sewage oxidation pond sample collected in New Zealand (Kraberger *et al.*, 2015a). Evidently, faecal sources are a powerful tool for the discovery of divergent CRESS DNA viruses in New Zealand.

In addition to faecal sources, CRESS DNA viruses have been recovered from other isolation sources in New Zealand. Additional gemycircularvirus isolates were identified in benthic river sediment (Kraberger *et al.*, 2013) and a soft brome sample (Kraberger *et al.*, 2015b). 46 *Beak and feather disease virus* isolates and two *Starling circovirus* isolates have been recovered from the blood and feathers of infected birds (Jackson *et al.*, 2014a; Jackson *et al.*, 2015; Massaro *et al.*, 2012; Ortiz-Catedral *et al.*, 2010) and mollusc tissue (Dayaram *et al.*, 2013a), respectively. The only cyclovirus to be identified in New Zealand, Dragonfly cyclovirus 7, was recovered from a red damselfly (Dayaram *et al.*, 2013c). Divergent, unclassified CRESS DNA viruses have also been identified in dragonflies (Dayaram *et al.*, 2014; Dayaram *et al.*, 2016), grass (Kraberger *et al.*, 2015b), molluscs (Dayaram *et al.*, 2016; Dayaram *et al.*, 2015a; Dayaram *et al.*, 2013b) and lake water and sediment (Dayaram *et al.*, 2016).

Given the limited research on this apparently prevalent group of viruses in a New Zealand context, this study aimed to identify novel CRESS DNA viruses. Additionally, a previous study identified a CRESS DNA virus, *Starling circovirus*, that was not known to be present in New Zealand using viral metagenomics (Dayaram *et al.*, 2013b). By identifying novel CRESS DNA viruses and previously identified CRESS DNA viruses from new isolation sources and locations, this study aims to expand the known diversity and distribution of this group of viruses.

Table 1.3: Description of circoviruses, cycloviruses, gemycircularviruses, smacoviruses, unclassified CRESS DNA viruses and unclassified circular molecules recovered from samples collected in New Zealand between 2008 and 2013.

CRESS virus type	Accession #	Virus	Host	Sample type	Sampling year	Sampling location	Reference
Circovirus	GQ396652	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Ortiz-Catedral *et al.*, 2010)
Circovirus	GQ396653	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Ortiz-Catedral *et al.*, 2010)
Circovirus	GQ396654	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Ortiz-Catedral *et al.*, 2010)
Circovirus	GQ396655	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Ortiz-Catedral *et al.*, 2010)
Circovirus	GQ396656	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Ortiz-Catedral *et al.*, 2010)
Circovirus	GU936287	BFDV	*Platycercus eximius*	Blood / feather	2008	Auckland	(Massaro *et al.*, 2012)
Circovirus	GU936288	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Massaro *et al.*, 2012)
Circovirus	GU936289	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Massaro *et al.*, 2012)
Circovirus	GU936290	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Massaro *et al.*, 2012)
Circovirus	GU936291	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Massaro *et al.*, 2012)
Circovirus	GU936292	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Massaro *et al.*, 2012)
Circovirus	GU936293	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Massaro *et al.*, 2012)
Circovirus	GU936294	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Massaro *et al.*, 2012)
Circovirus	GU936295	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Massaro *et al.*, 2012)
Circovirus	GU936296	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Massaro *et al.*, 2012)
Circovirus	GU936297	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2008	Little barrier Island	(Massaro *et al.*, 2012)
Circovirus	JF519618	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2010	Little barrier Island	(Massaro *et al.*, 2012)
Circovirus	JF519619	BFDV	*Platycercus eximius*	Blood / feather	2010	Auckland	(Massaro *et al.*, 2012)
Circovirus	JQ782196	BFDV	*Platycercus eximius*	Blood / feather	2009	Auckland	(Massaro *et al.*, 2012)
Circovirus	JQ782197	BFDV	*Platycercus eximius*	Blood / feather	2009	Auckland	(Massaro *et al.*, 2012)
Circovirus	JQ782198	BFDV	*Platycercus eximius*	Blood / feather	2009	Auckland	(Massaro *et al.*, 2012)
Circovirus	JQ782199	BFDV	*Platycercus eximius*	Blood / feather	2009	Auckland	(Massaro *et al.*, 2012)
Circovirus	JQ782200	BFDV	*Platycercus eximius*	Blood / feather	2009	Auckland	(Massaro *et al.*, 2012)
Circovirus	JQ782201	BFDV	*Cyanoramphus auriceps*	Blood / feather	2011	Fiordland	(Massaro *et al.*, 2012)
Circovirus	JQ782202	BFDV	*Cyanoramphus auriceps*	Blood / feather	2012	Fiordland	(Massaro *et al.*, 2012)
Circovirus	JQ782203	BFDV	*Cyanoramphus auriceps*	Blood / feather	2012	Fiordland	(Massaro *et al.*, 2012)
Circovirus	JQ782204	BFDV	*Cyanoramphus auriceps*	Blood / feather	2012	Fiordland	(Massaro *et al.*, 2012)
Circovirus	JQ782205	BFDV	*Cyanoramphus auriceps*	Blood / feather	2012	Fiordland	(Massaro *et al.*, 2012)
Circovirus	JQ782206	BFDV	*Cyanoramphus auriceps*	Blood / feather	2012	Fiordland	(Massaro *et al.*, 2012)
Circovirus	JQ782207	BFDV	*Cyanoramphus auriceps*	Blood / feather	2012	Fiordland	(Massaro *et al.*, 2012)
Circovirus	JQ782208	BFDV	*Cyanoramphus auriceps*	Blood / feather	2012	Fiordland	(Massaro *et al.*, 2012)
Circovirus	KC846095	StCV	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2013a)
Circovirus	KC846096	StCV	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2013a)
Circovirus	KF467251	BFDV	*Platycercus eximius*	Blood / feather	2012	Auckland	(Jackson *et al.*, 2014a)
Circovirus	KF467252	BFDV	*Platycercus eximius*	Blood / feather	2012	Auckland	(Jackson *et al.*, 2014a)
Circovirus	KF467253	BFDV	*Platycercus eximius*	Blood / feather	2012	Auckland	(Jackson *et al.*, 2014a)
Circovirus	KF467254	BFDV	*Platycercus eximius*	Blood / feather	2012	Auckland	(Jackson *et al.*, 2014a)
Circovirus	KM452734	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2013	Auckland	(Jackson *et al.*, 2015)
Circovirus	KM452735	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2013	Little barrier Island	(Jackson *et al.*, 2015)
Circovirus	KM452736	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2013	Little barrier Island	(Jackson *et al.*, 2015)
Circovirus	KM452737	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2011	Tiritiri Matangi Island	(Jackson *et al.*, 2015)
Circovirus	KM452738	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2011	Tiritiri Matangi Island	(Jackson *et al.*, 2015)

Circovirus	KM452739	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2012	Kapiti Island	(Jackson *et al.*, 2015)
Circovirus	KM452740	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2012	Kapiti Island	(Jackson *et al.*, 2015)
Circovirus	KM452741	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2012	Kapiti Island	(Jackson *et al.*, 2015)
Circovirus	KM452742	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2010	Kapiti Island	(Jackson *et al.*, 2015)
Circovirus	KM452743	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2010	Kapiti Island	(Jackson *et al.*, 2015)
Circovirus	KM452744	BFDV	*Cyanoramphus novaezelandiae*	Blood / feather	2009	ZEALANDIA-Karori sanctuary	(Jackson *et al.*, 2015)
Cyclovirus	KC512919	DfCyV-7	*Xanthocnemis zealandica*	Tissue	2011	Christchurch	(Dayaram *et al.*, 2013c)
Gemycircularvirus	KF268025	SsHADV-1	-	River Sediments	2012	Christchurch	(Kraberger *et al.*, 2013)
Gemycircularvirus	KF268026	SsHADV-1	-	River Sediments	2012	Christchurch	(Kraberger *et al.*, 2013)
Gemycircularvirus	KF268027	SsHADV-1	-	River Sediments	2012	Christchurch	(Kraberger *et al.*, 2013)
Gemycircularvirus	KF268028	SsHADV-1	-	River Sediments	2012	Christchurch	(Kraberger *et al.*, 2013)
Gemycircularvirus	KF371630	FaGmV-12	*Struthio camelus*	Faeces	2011	Christchurch	(Sikorski *et al.*, 2013d)
Gemycircularvirus	KF371631	FaGmV-11	*Oryctolagus cuniculus*	Faeces	2009	Cass Basin	(Sikorski *et al.*, 2013d)
Gemycircularvirus	KF371632	FaGmV-10	*Sturnus vulgaris*	Faeces	2009	Rangatira Island, Chatham Islands	(Sikorski *et al.*, 2013d)
Gemycircularvirus	KF371633	FaGmV-9	*Turdus merula*	Faeces	2011	Rangatira Island, Chatham Islands	(Sikorski *et al.*, 2013d)
Gemycircularvirus	KF371634	FaGmV-8	*Petroica traversi*	Faeces	2011	Rangatira Island, Chatham Islands	(Sikorski *et al.*, 2013d)
Gemycircularvirus	KF371635	FaGmV-7	*Anas platyrhynchos*	Faeces	2012	Christchurch	(Sikorski *et al.*, 2013d)
Gemycircularvirus	KF371636	FaGmV-6	*Gerygone albofrontata*	Faeces	2011	Rangatira Island, Chatham Islands	(Sikorski *et al.*, 2013d)
Gemycircularvirus	KF371637	FaGmV-5	*Gerygone albofrontata*	Faeces	2011	Rangatira Island, Chatham Islands	(Sikorski *et al.*, 2013d)
Gemycircularvirus	KF371638	FaGmV-4	*Arctocephalus forsteri*	Faeces	2012	Kaikoura	(Sikorski *et al.*, 2013d)
Gemycircularvirus	KF371639	FaGmV-3	*Gerygone albofrontata*	Faeces	2011	Rangatira Island, Chatham Islands	(Sikorski *et al.*, 2013d)
Gemycircularvirus	KF371640	FaGmV-2	*Sus scrofa*	Faeces	2011	Cass Basin	(Sikorski *et al.*, 2013d)
Gemycircularvirus	KF371641	FaGmV-1c	*Turdus merula*	Faeces	2009	Rangatira Island, Chatham Islands	(Sikorski *et al.*, 2013d)
Gemycircularvirus	KF371642	FaGmV-1b	*Turdus merula*	Faeces	2011	Rangatira Island, Chatham Islands	(Sikorski *et al.*, 2013d)
Gemycircularvirus	KF371643	FaGmV-1a	*Ovis aries*	Faeces	2009	Cass Basin	(Sikorski *et al.*, 2013d)
Gemycircularvirus	KJ547634	SaGmV-4	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Gemycircularvirus	KJ547635	SaGmV-5	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Gemycircularvirus	KJ547636	SaGmV-6	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Gemycircularvirus	KJ547637	SaGmV-7a	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Gemycircularvirus	KJ547638	SaGmV-8	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Gemycircularvirus	KJ547639	SaGmV-9	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Gemycircularvirus	KJ547640	SaGmV-7b	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Gemycircularvirus	KJ547641	SaGmV-11	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Gemycircularvirus	KJ547642	SaGmV-2	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Gemycircularvirus	KJ547643	SaGmV-3	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Gemycircularvirus	KJ547644	SaGmV-10a	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Gemycircularvirus	KJ547645	SaGmV-10b	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Gemycircularvirus	KM510192	BasCV-3	*Bromus hordeaceus*	Leaf	2012	Sefton	(Kraberger *et al.*, 2015b)
Gemycircularvirus	KM821747	SaGmV-1	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Smacovirus	JX274036	PoSCV	*Sus scrofa*	Faeces	2012	Cass Basin	(Sikorski *et al.*, 2013a)
Unclassified CRESS DNA virus	JX908739	CynNCXV	-	Nesting material	2012	Polter Valley	(Sikorski *et al.*, 2013c)
Unclassified CRESS DNA virus	JX908740	CynNCKV	-	Nesting material	2012	Polter Valley	(Sikorski *et al.*, 2013c)
Unclassified CRESS DNA virus	KC172652	GaCSV	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2013b)
Unclassified CRESS DNA virus	KF246569	FSfaCV	*Arctocephalus forsteri*	Faeces	2012	Kaikoura	(Sikorski *et al.*, 2013b)
Unclassified CRESS DNA virus	KF738873	DflaCV-1	*Procordulia grayi*	Tissue	2012	Lake Donne	(Dayaram *et al.*, 2014)
Unclassified CRESS DNA virus	KF738874	DflaCV-2	*Procordulia grayi*	Tissue	2012	Lake Sarah	(Dayaram *et al.*, 2014)
Unclassified CRESS DNA virus	KF738875	DflaCV-3	*Procordulia grayi*	Tissue	2012	Lake Grassmere	(Dayaram *et al.*, 2014)
Unclassified CRESS DNA virus	KF738876	DflaCV-3	*Procordulia grayi*	Tissue	2012	Lake Hawdon	(Dayaram *et al.*, 2014)

Unclassified CRESS DNA virus	KF738877	DflaCV-4	*Procordulia grayi*	Tissue	2012	Lake Grassmere	(Dayaram *et al.*, 2014)
Unclassified CRESS DNA virus	KF738878	DflaCV-5	*Procordulia grayi*	Tissue	2012	Lake Grassmere	(Dayaram *et al.*, 2014)
Unclassified CRESS DNA virus	KF738879	DflaCV-5	*Procordulia grayi*	Tissue	2012	Lake Sarah	(Dayaram *et al.*, 2014)
Unclassified CRESS DNA virus	KF738880	DflaCV-6	*Procordulia grayi*	Tissue	2012	Lake Donne	(Dayaram *et al.*, 2014)
Unclassified CRESS DNA virus	KF738881	DflaCV-7	*Procordulia grayi*	Tissue	2012	Lake Hawdon	(Dayaram *et al.*, 2014)
Unclassified CRESS DNA virus	KF738882	DflaCV-8	*Procordulia grayi*	Tissue	2012	Lake Sarah	(Dayaram *et al.*, 2014)
Unclassified CRESS DNA virus	KF738883	DflaCV-9	*Procordulia grayi*	Tissue	2012	Lake Donne	(Dayaram *et al.*, 2014)
Unclassified CRESS DNA virus	KF738884	DflaCV-10	*Xanthocnemis zealandica*	Tissue	2012	Lake Sarah	(Dayaram *et al.*, 2014)
Unclassified CRESS DNA virus	KF738885	DflaCV-10	*Xanthocnemis zealandica*	Tissue	2012	Lake Hawdon	(Dayaram *et al.*, 2014)
Unclassified CRESS DNA virus	KJ547620	SaCV-1	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KJ547621	SaCV-10	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KJ547622	SaCV-11	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KJ547623	SaCV-12	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KJ547624	SaCV-13	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KJ547625	SaCV-14	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KJ547626	SaCV-2	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KJ547627	SaCV-3	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KJ547628	SaCV-4	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KJ547629	SaCV-5	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KJ547630	SaCV-6	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KJ547631	SaCV-7	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KJ547632	SaCV-8	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KJ547633	SaCV-9	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM510189	BasCV-1	*Bromus hordeaceus*	Leaf	2012	Sefton	(Kraberger *et al.*, 2015b)
Unclassified CRESS DNA virus	KM510190	BasCV-1	*Bromus hordeaceus*	Leaf	2012	Wellington	(Kraberger *et al.*, 2015b)
Unclassified CRESS DNA virus	KM510191	BasCV-2	*Bromus hordeaceus*	Leaf	2012	Wellington	(Kraberger *et al.*, 2015b)
Unclassified CRESS DNA virus	KM821748	SaCV-36	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821749	SaCV-37	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821750	SaCV-15	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821751	SaCV-16	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821752	SaCV-17	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821753	SaCV-18	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821754	SaCV-19	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821755	SaCV-20	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821756	SaCV-21	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821757	SaCV-22	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821758	SaCV-23	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821759	SaCV-24	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821760	SaCV-25	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821761	SaCV-26	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821762	SaCV-27	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821763	SaCV-28	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821764	SaCV-29	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821765	SaCV-30	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821766	SaCV-31	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821767	SaCV-32	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821768	SaCV-33	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM821769	SaCV-34	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)

Unclassified CRESS DNA virus	KM821770	SaCV-35	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874290	AHEaCV-1	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874291	AHEaCV-1	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874292	AHEaCV-2	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874293	AHEaCV-2	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874294	AHEaCV-2	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874295	AHEaCV-3	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874296	AHEaCV-3	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874297	AHEaCV-3	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874298	AHEaCV-4	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874299	AHEaCV-4	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874300	AHEaCV-4	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874301	AHEaCV-5	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874302	AHEaCV-5	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874303	AHEaCV-5	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874304	AHEaCV-6	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874305	AHEaCV-6	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874306	AHEaCV-6	-	River Sediments	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874307	AHEaCV-6	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874308	AHEaCV-6	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874309	AHEaCV-7	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874310	AHEaCV-8	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874311	AHEaCV-8	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874312	AHEaCV-8	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874313	AHEaCV-8	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874314	AHEaCV-8	-	River Sediments	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874315	AHEaCV-9	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874316	AHEaCV-9	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874317	AHEaCV-9	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874318	AHEaCV-9	-	River Sediments	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874319	AHEaCV-10	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874320	AHEaCV-10	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874321	AHEaCV-10	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874322	AHEaCV-10	-	River Sediments	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874323	AHEaCV-11	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874324	AHEaCV-11	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874325	AHEaCV-11	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874326	AHEaCV-11	-	River Sediments	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874327	AHEaCV-11	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874328	AHEaCV-12	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874329	AHEaCV-13	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874330	AHEaCV-13	-	River Sediments	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874331	AHEaCV-13	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874332	AHEaCV-14	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874333	AHEaCV-14	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874334	AHEaCV-14	-	River Sediments	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874335	AHEaCV-14	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874336	AHEaCV-15	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)

Unclassified CRESS DNA virus	KM874337	AHEaCV-15	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874338	AHEaCV-15	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874339	AHEaCV-15	-	River Sediments	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874340	AHEaCV-16	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874341	AHEaCV-16	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874342	AHEaCV-16	-	River Sediments	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874343	AHEaCV-17	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874344	AHEaCV-17	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874345	AHEaCV-17	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874346	AHEaCV-18	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874347	AHEaCV-19	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874348	AHEaCV-20	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874349	AHEaCV-20	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874350	AHEaCV-21	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874351	AHEaCV-22	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874352	AHEaCV-22	-	River Sediments	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874353	AHEaCV-23	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874354	AHEaCV-24	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874355	AHEaCV-25	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874356	AHEaCV-25	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874357	AHEaCV-25	-	River Sediments	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874358	AHEaCV-25	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874359	AHEaCV-26	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874360	AHEaCV-27	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874361	AHEaCV-27	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874362	AHEaCV-28	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874363	AHEaCV-28	*Amphibola crenata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874364	AHEaCV-28	-	River Sediments	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874365	AHEaCV-28	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874366	AHEaCV-29	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874367	AHEaCV-29	*Austrovenus stutchburyi*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KM874368	AHEaCV-29	*Paphies subtriangulata*	Tissue	2012	Christchurch	(Dayaram *et al.*, 2015a)
Unclassified CRESS DNA virus	KP153390	LSaCV-1	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al.*, 2016)
Unclassified CRESS DNA virus	KP153391	LSaCV-1	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al.*, 2016)
Unclassified CRESS DNA virus	KP153392	LSaCV-1	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al.*, 2016)
Unclassified CRESS DNA virus	KP153393	LSaCV-1	-	Water	2013	Lake Sarah	(Dayaram *et al.*, 2016)
Unclassified CRESS DNA virus	KP153394	LSaCV-2	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al.*, 2016)
Unclassified CRESS DNA virus	KP153395	LSaCV-3	*Chironomus zealandicus*	Tissue			
Unclassified CRESS DNA virus	KP153396	LSaCV-3	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al.*, 2016)
Unclassified CRESS DNA virus	KP153397	LSaCV-4	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al.*, 2016)
Unclassified CRESS DNA virus	KP153398	LSaCV-4	-	Sediment	2013	Lake Sarah	(Dayaram *et al.*, 2016)
Unclassified CRESS DNA virus	KP153399	LSaCV-4	-	Water			
Unclassified CRESS DNA virus	KP153400	LSaCV-4	*Echyridella menziesi*	Tissue			
Unclassified CRESS DNA virus	KP153401	LSaCV-4	*Procordulia grayi, Xanthocnemis zealandica*	Tissue	2013	Lake Sarah	(Dayaram *et al.*, 2016)
Unclassified CRESS DNA virus	KP153402	LSaCV-5	-	Sediment			
Unclassified CRESS DNA virus	KP153403	LSaCV-6	-	Sediment	2013	Lake Sarah	(Dayaram *et al.*, 2016)
Unclassified CRESS DNA virus	KP153404	LSaCV-7	-	Sediment	2013	Lake Sarah	(Dayaram *et al.*, 2016)
Unclassified CRESS DNA virus	KP153405	LSaCV-8	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al.*, 2016)

Unclassified CRESS DNA virus	KP153406	LSaCV-8	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153407	LSaCV-9	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153408	LSaCV-10	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153409	LSaCV-11	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153410	LSaCV-12	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153411	LSaCV-12	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153412	LSaCV-13	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153413	LSaCV-13	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153414	LSaCV-14	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153415	LSaCV-14	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153416	LSaCV-14	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153417	LSaCV-15	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153418	LSaCV-15	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153419	LSaCV-15	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153420	LSaCV-16	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153421	LSaCV-16	-	Water	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153422	LSaCV-16	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153423	LSaCV-16	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153424	LSaCV-17	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153425	LSaCV-17	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153426	LSaCV-17	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153427	LSaCV-18	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153428	LSaCV-18	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153429	LSaCV-19	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153430	LSaCV-19	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153431	LSaCV-19	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153432	LSaCV-19	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153433	LSaCV-19	*Procordulia grayi, Xanthocnemis zealandica*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153434	LSaCV-20	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153435	LSaCV-20	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153436	LSaCV-20	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153437	LSaCV-21	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153438	LSaCV-21	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153439	LSaCV-21	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153440	LSaCV-21	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153441	LSaCV-22	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153442	LSaCV-23	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153443	LSaCV-24	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153444	LSaCV-24	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153445	LSaCV-25	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153446	LSaCV-26	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153448	DflaCV-3	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153447	DflaCV-3	*Procordulia grayi, Xanthocnemis zealandica*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153449	DflaCV-3	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153450	DflaCV-3	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153451	LSaCV-28	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153452	LSaCV-28	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153453	LSaCV-28	*Procordulia grayi, Xanthocnemis zealandica*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)

Unclassified CRESS DNA virus	KP153454	LSaCV-29	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153455	LSaCV-29	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153456	LSaCV-30	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153457	LSaCV-30	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153458	LSaCV-30	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153459	LSaCV-31	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153460	LSaCV-31	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153461	LSaCV-31	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153462	LSaCV-31	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153463	LSaCV-31	*Procordulia grayi, Xanthocnemis zealandica*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153464	LSaCV-32	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153465	LSaCV-32	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153466	LSaCV-32	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153467	LSaCV-32	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153468	LSaCV-32	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153469	LSaCV-33	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153470	LSaCV-34	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153471	LSaCV-34	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153474	LSaCV-36	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153475	LSaCV-36	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153476	LSaCV-37	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153477	LSaCV-37	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153478	LSaCV-37	*Procordulia grayi, Xanthocnemis zealandica*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153479	LSaCV-37	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153480	LSaCV-37	-	Water	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153481	LSaCV-37	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153482	LSaCV-37	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153483	LSaCV-38	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153484	LSaCV-38	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153485	LSaCV-39	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153486	LSaCV-39	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153487	LSaCV-39	*Procordulia grayi, Xanthocnemis zealandica*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153488	LSaCV-40	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153489	LSaCV-40	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153490	LSaCV-35	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153491	LSaCV-35	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153492	LSaCV-35	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153493	LSaCV-35	*Procordulia grayi, Xanthocnemis zealandica*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153494	LSaCV-41	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153495	LSaCV-41	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153496	LSaCV-42	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153497	LSaCV-42	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153498	LSaCV-42	*Procordulia grayi, Xanthocnemis zealandica*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153499	LSaCV-43	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153500	LSaCV-44	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153501	LSaCV-45	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153502	LSaCV-46	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153503	LSaCV-46	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)

Unclassified CRESS DNA virus	KP153504	LSaCV-47	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153505	LSaCV-48	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153506	LSaCV-48	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153507	LSaCV-49	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153508	LSaCV-49	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153509	LSaCV-49	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153510	LSaCV-49	-	Water	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153511	LSaCV-50	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153512	LSaCV-50	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153513	LSaCV-50	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153514	LSaCV-50	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153515	LSaCV-50	*Procordulia grayi, Xanthocnemis zealandica*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153516	DflaCV-10	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153517	DflaCV-10	-	Water	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153518	DflaCV-10	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153519	DflaCV-10	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153520	DflaCV-10	*Procordulia grayi, Xanthocnemis zealandica*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153521	DflaCV-10	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153522	LSaCV-51	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153523	LSaCV-27	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153524	DflaCV-5	-	Water	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153525	DflaCV-5	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153526	DflaCV-8	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified CRESS DNA virus	KP153527	DflaCV-6	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KJ547617	SaCM-2	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified circular molecule	KJ547618	SaCM-1	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified circular molecule	KJ547619	SaCM-3	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified circular molecule	KM877826	SaCM-4	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified circular molecule	KM877827	SaCM-5	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified circular molecule	KM877828	SaCM-6	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified circular molecule	KM877829	SaCM-7	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified circular molecule	KM877830	SaCM-8	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified circular molecule	KM877831	SaCM-9	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified circular molecule	KM877832	SaCM-10	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified circular molecule	KM877833	SaCM-11	-	Sewage	2012	Christchurch	(Kraberger *et al.*, 2015a)
Unclassified circular molecule	KP153359	LSaCM-1	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153360	LSaCM-2	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153361	LSaCM-3	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153362	LSaCM-3	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153363	LSaCM-4	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153364	LSaCM-5	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153365	LSaCM-5	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153366	LSaCM-5	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153367	LSaCM-5	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153368	LSaCM-6	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153369	LSaCM-6	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153370	LSaCM-7	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153371	LSaCM-8	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)

Unclassified circular molecule	KP153372	LSaCM-8	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153373	LSaCM-8	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153374	LSaCM-8	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153375	LSaCM-8	-	Water	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153376	LSaCM-8	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153377	LSaCM-9	*Procordulia grayi, Xanthocnemis zealandica*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153378	LSaCM-9	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153379	LSaCM-10	*Procordulia grayi, Xanthocnemis zealandica*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153380	LSaCM-10	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153381	LSaCM-10	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153382	LSaCM-10	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153383	LSaCM-10	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153384	LSaCM-11	-	Water	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153385	LSaCM-11	*Chironomus zealandicus*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153386	LSaCM-11	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153387	LSaCM-11	*Echyridella menziesi*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153388	LSaCM-11	-	Sediment	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153389	LSaCM-11	*Musculium novazelandiae*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153472	LSaCM-12	*Potamopyrgus antipodarum, Physella acuta*	Tissue	2013	Lake Sarah	(Dayaram *et al*., 2016)
Unclassified circular molecule	KP153473	LSaCM-12	-	Water	2013	Lake Sarah	(Dayaram *et al*., 2016)

AHEaCV: Avon-Heathcote Estuary associated circular virus
BFDV: Beak and feather disease virus
BasCV: Bromus-associated circular DNA virus
CynNCKV: Cyanoramphus nest-associated circular K DNA virus
CynNCXV: Cyanoramphus nest-associated circular X DNA virus
DfCyV: Dragonfly cyclovirus
DflaCV: Dragonfly larvae associated circular virus
FaGmV: Faecal-associated gemycircularvirus
FSfaCV: Fur seal feces associated circular DNA virus
GasCSV: Gastropod-associated circular ssDNA virus
LSaCM: Lake Sarah-associated circular DNA molecule
LSaCV: Lake Sarah-associated circular DNA virus
PoSCV: Porcine stool-associated circular virus
SsHADV-1: *Sclerotinia sclerotiorum* hypovirulence associated DNA virus
SaCM: Sewage-associated circular DNA molecule
SaCV: Sewage-associated circular DNA virus
SaGmV: Sewage-associated gemycircularvirus
StCV: Starling circovirus

Chapter Two: Isolation and characterisation of CRESS DNA viruses and circular molecules from animal faecal matter

Table of Contents

2.1 Abstract

In recent years, innovations in molecular techniques and sequencing technologies have resulted in a rapid expansion in the number of known viral sequences, particularly of CRESS DNA viruses. CRESS DNA viruses are virome components of many ecosystems and are known to infect a wide range of organisms. A large number of CRESS DNA viruses cannot be classified into viral taxa, indicating that the current view of the CRESS DNA viral sequence space is greatly underestimated. Animal faecal matter has proven to be a particularly useful source for sampling CRESS DNA viruses in an ecosystem, as it is cost-effective, non-invasive and harbours many potential host species. In this study, a viral metagenomic approach was used to explore the diversity of CRESS DNA viruses present in the faeces of domestic and wild animals in New Zealand. From 49 individual animal faecal samples, 38 complete CRESS DNA viral genomes and two circular molecules that may be defective molecules or single components of multicomponent genomes were identified. Based on shared genome organisations and sequence similarities, eighteen of the isolates were classified as gemycircularviruses and twelve isolates were classified as smacoviruses. The remaining eight isolates lack significant sequence similarity with any members of known CRESS DNA virus groupings. This research adds significantly to our knowledge of CRESS DNA virus diversity in New Zealand and emphasises the prevalence of CRESS DNA viruses in nature.

2.2 Introduction

Over the last decade there has been a significant increase in our knowledge of viruses not only in diseased organisms but also in non-symptomatic organisms and environmental samples through viral metagenomic approaches. Next-generation sequencing-based viral metagenomic approaches have enabled the identification of viruses that are unable to be cultured using traditional methods (Breitbart *et al.*, 2003). Furthermore, rolling circle amplification has enabled the rapid discovery of small circular DNA viruses (Johne *et al.*, 2009). As a result there has been a significant increase in both known and novel viral sequences of circular dsDNA viruses (papillomaviruses, polyomaviruses and hepadnaviruses) and ssDNA viruses (anelloviruses, circoviruses, nanoviruses and geminiviruses) that are known to infect eukaryotes. Additionally, a large number of CRESS DNA viruses have been discovered in a range of sample types. The Reps of CRESS DNA viruses share similarities to those of circoviruses, nanoviruses and geminiviruses (Rosario *et al.*, 2012b).

Circoviridae, Geminiviridae and *Nanoviridae* are the only three CRESS DNA viral families that are currently recognised by the ICTV (King *et al*., 2011; Rosario *et al*., 2012b). The Rep is reasonably well conserved across these three families compared to the CP. Since the Rep initiates RCR and has helicase activity, conserved motifs can be identified within the Reps of these families (Rosario *et al*., 2012b). The N-termini of Reps contain the RCR motifs that are involved in initiating the RCR mechanism. RCR motif I is crucial in dsDNA binding, RCR motif II coordinates divalent metal ions through two highly conserved histadine residues, and RCR motif III is proposed to mediate endonuclease activity using coordinated metal ions and invariant tyrosine and lysine residues, as reviewed in Rosario *et al.* (2012b). A fourth conserved motif, the geminivirus Rep sequence (GRS), mediates appropriate spatial arrangements of RCR motifs II and III and is only found in geminiviruses (Nash *et al.*, 2011) and gemycircularviruses (Dayaram *et al.*, 2012; Rosario *et al.*, 2012a). Rep helicase activity is mediated by a C-terminal NTP-binding domain in which geminivirus Reps contain superfamily 3 (SF3) helicase domains, namely Walker A, Walker B and motif C (Rosario *et al.*, 2012b).

Identification of viruses through sampling of animal faecal matter has proved to be a very useful approach for the discovery of a wide variety of viral types, in particular CRESS DNA

viruses and small dsDNA viruses (Blinkova *et al.*, 2010; Castrignano *et al.*, 2013; Cheung *et al.*, 2014a; Cheung *et al.*, 2014b; Cheung *et al.*, 2013; Cheung *et al.*, 2015; Conceicao-Neto *et al.*, 2015; Delwart & Li, 2012; Ge *et al.*, 2012; Hansen *et al.*, 2015; Kim *et al.*, 2014; Kim *et al.*, 2012; Kraberger *et al.*, 2015a; Li *et al.*, 2010a; Li *et al.*, 2011; Li *et al.*, 2010b; Li *et al.*, 2015; Male *et al.*, 2016; Ng *et al.*, 2014; Ng *et al.*, 2012; Phan *et al.*, 2015; Reuter *et al.*, 2014; Sachsenröder *et al.*, 2014; Sachsenröder *et al.*, 2012; Sasaki *et al.*, 2015; Shan *et al.*, 2011; Sikorski *et al.*, 2013a; Sikorski *et al.*, 2013b; Sikorski *et al.*, 2013d; van den Brand *et al.*, 2012; Varsani *et al.*, 2014a; Varsani *et al.*, 2015; Victoria *et al.*, 2009; Woo *et al.*, 2014; Wu *et al.*, 2015; Zhang *et al.*, 2014). The major advantage of faecal sampling for viruses is that it is non-invasive to the animal. In addition to faecal matter, CRESS DNA viruses have also been identified in leaf samples (Basso *et al.*, 2015; Dayaram *et al.*, 2012; Du *et al.*, 2014; Kraberger *et al.*, 2015b; Male *et al.*, 2015; Marzano & Domier, 2015), adult insects (Dayaram *et al.*, 2014; Dayaram *et al.*, 2013c; Dayaram *et al.*, 2015b; Garigliany *et al.*, 2015; Ng *et al.*, 2011b; Padilla-Rodriguez *et al.*, 2013; Pham *et al.*, 2013; Rosario *et al.*, 2012a; Rosario *et al.*, 2011), air samples (Roux *et al.*, 2013; Whon *et al.*, 2012), fresh water ecosystems (Dayaram *et al.*, 2014; Dayaram *et al.*, 2016; Hewson *et al.*, 2013a; Hewson *et al.*, 2013b; Roux *et al.*, 2012; Smith *et al.*, 2013; Zawar-Reza *et al.*, 2014), marine ecosystems (Angly *et al.*, 2006; Breitbart *et al.*, 2015; Dayaram *et al.*, 2015a; Dunlap *et al.*, 2013; Fahsbender *et al.*, 2015; Labonté & Suttle, 2013; Ng *et al.*, 2013; Rosario *et al*., 2009a; Rosario *et al.*, 2009b; Rosario *et al.*, 2015a; Soffer *et al.*, 2014; Yoshida *et al.*, 2013), soil (Kim *et al.*, 2008; Reavy *et al.*, 2015) and wastewater / sewage (Cantalupo *et al.*, 2011; Kraberger *et al.*, 2015a; Ng *et al.*, 2012; Phan *et al.*, 2015; Rosario *et al.*, 2009; Roux *et al.*, 2013).

A number of CRESS DNA viruses cannot be formally classified into pre-existing viral taxa. Clusters of unclassified CRESS DNA viruses sharing similar sequences and genome organisations have resulted in the proposal of large, novel CRESS DNA viral groups. The number of viruses in these groupings has expanded rapidly in recent years. Studies suggest that these groups are diverse and have a broad distribution.

One of the proposed groupings is the gemycircularviruses (Rosario *et al.*, 2012a). Gemycircularviruses encode a Rep with similarities to that of the geminiviruses (Yu *et al.*, 2010). The first gemycircularvirus described, Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1 (SsHADV-1), was identified from *Sclerotinia sclerotiorum* isolates exhibiting slowed growth and morphology indicative of hypovirulence (Yu *et al.*, 2010).

SsHADV-1 isolates have subsequently been recovered from benthic river sediments in New Zealand (Kraberger *et al.*, 2013) and dragonfly samples collected in the United States of America (Dayaram *et al.*, 2015b) suggesting that SsHADV-1 has a wider global distribution. Gemycircularviruses have been identified in insects (Dayaram *et al.*, 2015b; Ng *et al.*, 2011b; Rosario *et al.*, 2012a), animal blood, serum and cerebrospinal fluid (Lamberto *et al.*, 2014; Li *et al.*, 2015; Phan *et al.*, 2015; Uch *et al.*, 2015), animal cervical, cloacal, pharyngeal and rectal swabs (Hanna *et al.*, 2015; van den Brand *et al.*, 2012; Wu *et al.*, 2015), animal and human faecal matter (Conceicao-Neto *et al.*, 2015; Male *et al.*, 2016; Ng *et al.*, 2014; Phan *et al.*, 2015; Sikorski *et al.*, 2013d), sewage (Kraberger *et al.*, 2015a; Phan *et al.*, 2015) and plant material (Dayaram *et al.*, 2012; Du *et al.*, 2014; Kraberger *et al.*, 2015b; Male *et al.*, 2015; Marzano & Domier, 2015) from various geographical regions.

Another proposed grouping is the smacoviruses. The Reps encoded by smacoviruses are phylogenetically distinct from other CRESS DNA virus families and display a high level of sequence diversity (Ng *et al.*, 2015). Accordingly, Ng *et al.* (2015) classified smacovirus isolates into one of eleven genogroups. The majority of smacoviruses have been identified in human and animal faeces (Blinkova *et al.*, 2010; Cheung *et al.*, 2013; Cheung *et al.*, 2015; Kim *et al.*, 2014; Kim *et al.*, 2012; Ng *et al.*, 2015; Reuter *et al.*, 2014; Sachsenröder *et al.*, 2014; Sachsenröder *et al.*, 2012; Sikorski *et al.*, 2013a; Woo *et al.*, 2014). Additional smacoviruses have been identified from a dragonfly sample (Dayaram *et al.*, 2015b) and sewage oxidation pond sample (Kraberger *et al.*, 2015a).

In addition, a new *Cyclovirus* genus (*Circoviridae* family) has recently emerged (Li *et al.*, 2011). Cycloviruses encode a Rep most closely related to that of the circoviruses, however, cycloviruses cluster within a monophyletic clade, distinct from circoviruses. Cycloviruses have been identified in samples collected from dragonflies (Dayaram *et al.*, 2013c; Rosario *et al.*, 2012a; Rosario *et al.*, 2011) and a variety of animals (Ge *et al.*, 2011; Li *et al.*, 2011; Li *et al.*, 2015; Lima *et al.*, 2015; Male *et al.*, 2016; Sasaki *et al.*, 2015; Sato *et al.*, 2015; Tan *et al.*, 2013; Wu *et al.*, 2015; Zawar-Reza *et al.*, 2014).

In a bid to further explore the CRESS DNA viral diversity associated with domestic and wild animals in New Zealand, a viral metagenomics study on their faecal matter was conducted. This study identified 38 CRESS DNA viruses and two circular molecules from faecal samples of *Anas platyrhynchos*, *Bos taurus*, *Canis lupus familiaris*, *Dama dama*, *Equus ferus*,

Gallus gallus domesticus, *Lama glama*, *Lepus europaeus*, *Ovis aries*, *Rupicapra rupicapra*, *Sus scrofa domesticus*, *Trichosurus vulpecula*. All samples were collected on the South Island of New Zealand.

2.3 Materials and methods

2.3.1 Sample preparation, circular DNA enrichment, Illumina sequencing and viral genome recovery

Faecal samples of 49 wild and domestic animals were collected between 2009 and 2013 from various sites across the South Island of New Zealand. Approximately five grams of the faecal sample was resuspended in SM buffer [0.1M NaCl, 50mM Tris-HCl (pH 7.4), 10mM $MgSO_4$] and homogenized. The homogenate was centrifuged at 10000 x g for ten minutes and filtered sequentially through 0.45µm and 0.2µm syringe filters (Sartorius Stedim Biotech, Germany). The viral particles were precipitated using 15% polyethylene glycol (PEG). Following this, the solution was centrifuged at 10000 x g for 10 minutes. The pellet was resuspended in 500µl of SM buffer and viral DNA was extracted from 200µl of resuspension using the High Pure Viral Nucleic Acid Kit (Roche Diagnostics, USA), according to the manufacturer's instructions.

As this study targeted CRESS DNA viruses, TempliPhi^{TM} (GE Healthcare, USA) was used to enrich circular DNA using rolling circle amplification. Enriched DNA from all samples was pooled and sequenced on an Illumina HiSeq 2000 sequencer at Beijing Genomics Institute (Hong Kong). The resulting pair-end reads were *de novo* assembled (kmer=64) using ABYSS 1.5.2 (Simpson *et al.*, 2009). Contigs >500 nts (n=708) were analysed by BLASTx against a viral protein database. Of these, 222 contigs were found to have viral hits to CRESS DNA viral proteins. Abutting primers were designed based on the contigs to recover complete CRESS DNA viral genomes and circular molecules (Table 2.1). These primer pairs were used to screen and recover viral genomes from the faecal samples using KAPA HiFi Hotstart DNA polymerase (Kapa Biosystems, USA). The amplicons were resolved on a 0.7% agarose

Table 2.1: Primer pairs used to recover complete genomes of CRESS DNA viruses and circular molecules from the faecal matter of wild and domestic animals collected across New Zealand.

Accession #	Virus / molecule	Faecal source	Faecal source common name	Forward primer	Reverse primer
KT862218	BofSmV-2	*Bos taurus*	Cow	TGCACCGCATCTAGACCAGACTTCTTC	TCTTATCTGCCAGTGCTCATAGCCGC
KT862219	ShfSmV-3	*Ovis aries*	Sheep	TAATACCCTTGATGCCTCGAATACAGGAATC	CCGATGATGTCATAGCCTGCAGAATTG
KT862220	ShfSmV-1	*Ovis aries*	Sheep	GCATCATGGAAATACCCCTGAAGCCC	CTGAGCAGACCGGTGTTCCCTC
KT862221	ShfSmV-2	*Ovis aries*	Sheep	TTGATCGCGATCCGTACCCCTGGTAT	ACCCATCTTGTTCAGAGTGGTTGAAAGATC
KT862222	BofSmV-3	*Bos taurus*	Cow	CAGGAGTTTGTAAGCGGGTTTC	TCGGCGGATAGATGGGACGT
KT862223	BofSmV-1	*Bos taurus*	Cow	GTATCCTCTCCTGTCTGACATCGGTAATG	ATGAGAGGACGGAGGTTCCTCATGAC
KT862224	BofSmV-5	*Bos taurus*	Cow	GTAAAGACAGGCTCAGGAACGG	TCCTCAGTACATCCCTAAGTC
KT862225	PofSmV-1	*Sus scrofa domesticus*	Pig	CTGTTGAGTGTACGCCCTC	GAGGCATAGGAACAGGGC
KT862226	PoSCV	*Lepus europaeus*	Hare	AGGTTTCGTTCGAGGCTGGTG	GGAGAGGATCTGCGGGAAG
KT862227	PoSCV	*Trichosurus vulpecula*	Common brushtail possum	AGGTTTCGTTCGAGGCTGGTG	GGAGAGGATCTGCGGGAAG
KT862228	BofSmV-4	*Bos taurus*	Cow	TACCTCCGGCACAGATCCAAAC	ATGTAGTAGCTCACGCTTGCTCATC
KT862229	BofSmV-6	*Bos taurus*	Cow	CTCCTGGAGCTGAAGACGAC	AAGCGATGATAGGATCGACGTTGAATG
KT862230	BofCV-1	*Bos taurus*	Cow	TACAGATACCTTGAGTACAGGCCG	TTCCCTGAGTTCGCTTTCCCAC
KT862231	DufCV-1	*Anas platyrhynchos*	Duck	GAACAAGCTACAAAGAAGAAGTGT	CATGAACAGGTTGAGATCGAAC
KT862232	DefCV-1	*Dama dama*	Deer	AGGAACTACAATGGCGTCGCCGC	GAGCATCCACTTGGAGTGCTCTGTC
KT862233	DufCV-2	*Anas platyrhynchos*	Duck	ATATTCACTAGTCCACAACATTTCAATGATG	GATTGTTTCAGCTAACCATGGTATTCTAC
KT862234	DufCV-3	*Anas platyrhynchos*	Duck	ATATTCACTAGTCCACAACATTTCAATGATG	GATTGTTTCAGCTAACCATGGTATTCTAC
KT862235	LlfCV-1	*Lama glama*	Llama	ATGTCGAAGAGGAAGATGTG	CGTCTAACACCTCTTCGACATAGTTG
KT862236	ChfCV-1	*Rupicapra rupicapra*	Chamois	CGTCTAACACCTCTTCGACCA	ATTCGCCGAATCCAACTCC
KT862237	BofCV-2	*Bos taurus*	Cow	TTAAGATACCGCAACTCCAAAAG	AAATCGATACCATCCCTGATTCC
KT862238	FaGmV-14	*Anas platyrhynchos*	Duck	CAATTACTCGAGAGCTGGCACC	CCCAAAACTCGTCAACACTTG
KT862239	FaGmV-14	*Anas platyrhynchos*	Duck	CAATTACTCGAGAGCTGGCACC	CCCAAAACTCGTCAACACTTG
KT862240	SaGmV-3	*Gallus gallus domesticus*	Chicken	ACAGAAGTGCCCTTGGTG	GTTCATTCACCTCACTCCG
KT862241	FaGmV-4	*Gallus gallus domesticus*	Chicken	TGTCTTCGTGGAGCATTATCGT	CATCTTCGTAAACTCCTCTTCC
KT862242	FaGmV-17	*Gallus gallus domesticus*	Chicken	TGATGACCCTGCACATCAAAG	CCCAAACATCGAGCGATCTA
KT862243	FaGmV-20	*Gallus gallus domesticus*	Chicken	TGGTGACCATCAACATCGAAG	CCCAAACATTTCGCAGTCTCG
KT862244	FaGmV-20	*Lama glama*	Llama	TGGTGACCATCAACATCGAAG	CCCAAACATTTCGCAGTCTCG
KT862245	FaGmV-21	*Lama glama*	Llama	TGGTAGCCGAGTACATCGAAA	CCCAAACATTGAACCTTCTCG
KT862246	FaGmV-20	*Equus ferus*	Horse	TGGTGACCATCAACATCGAAG	CCCAAACATTTCGCAGTCTCG
KT862247	FaGmV-21	*Equus ferus*	Horse	TGGTAGCCGAGTACATCGAAA	CCCAAACATTGAACCTTCTCG
KT862248	FaGmV-18	*Equus ferus*	Horse	TGCTTGCCTCCGACATCAAA	CCCGAACGTTGTGCCATC
KT862249	FaGmV-16	*Ovis aries*	Sheep	TGCCTACCCTCCACATCGAA	CCCAAACGTTGTCCCTTCTAA
KT862250	FaGmV-19	*Sus scrofa domesticus*	Pig	TGCTTGCCTCCGACATCAAA	CCCGAACGTTGTGCCATC
KT862251	FaGmV-16	*Ovis aries*	Sheep	TGCCTACCCTCCACATCGAA	CCCAAACGTTGTCCCTTCTAA
KT862252	SaGmV-3	*Bos taurus*	Cow	ACAGAAGTGCCCTTGGTG	GTTCATTCACCTCACTCCG
KT862253	FaGmV-22	*Bos taurus*	Cow	TGACGACCGTCGACATCAAAG	CCCAAACATCGCGCCATCTT
KT862254	FaGmV-15	*Canis lupus familiaris*	Dog	CAATTACTCGAGAGCTGGCACC	CCCAAAACTCGTCAACACTTG
KT862255	SaGmV-3	*Lepus europaeus*	Hare	ACAGAAGTGCCCTTGGTG	GTTCATTCACCTCACTCCG
KT862256	PofCM-1	*Sus scrofa domesticus*	Pig	TCCCTTGGTCTCATAACGTCCTGG	GGATCAGTACTTATTTACGGACTCAAGAAG
KT862257	BofCM-1	*Bos taurus*	Cow	CGTCTAACACCTCTTCGACCA	ATTCGCCGAATCCAACCCC

gel, excised and purified using MEGAquick-spin^{TM} total fragment DNA purification kit (iNtRON Biotechnology, South Korea). The gel-purified amplicons were ligated into a pJET1.2 vector (Thermo Fisher Scientific, USA) and recombinant plasmids were transformed into competent *Escherichia coli* DH5α cells. Plasmid DNA was purified from positive transformants using the DNA-spin plasmid DNA purification kit (iNtRON Biotechnology, South Korea) and the purified plasmids were sequenced at Macrogen Inc. (South Korea) by primer walking.

2.3.2 Sequence analysis

The complete genomes were assembled from reads generated by Sanger sequencing using DNA Baser V4 (Heracle Biosoft S.R.L., Romania). ORFs were identified using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and BLASTx was used to confirm ORF identity in GenBank's non-redundant database (Altschul *et al.*, 1990). Pairwise identities of the full genomes and CP and Rep sequences of CRESS DNA viruses / molecules were determined using Sequence Demarcation Tool (SDT) version 1.2 (Muhire *et al.*, 2014).

Genome sequences of all gemycircularviruses (Table 2.2) and smacoviruses (Table 2.3) were downloaded from GenBank on the 5th of January 2016. The Rep and CP amino acid sequences of gemycircularviruses and smacoviruses, including those from this study, were aligned with PROMALS3D (Pie *et al.*, 2008). These alignments were used to infer maximum-likelihood phylogenetic trees using PHYML (Guindon *et al.*, 2010) and best fit substitution models were determined using ProtTest (Darriba *et al.*, 2011) with approximate likelihood branch support (aLRT). The VT+I+G and LG+I+G models were used for the gemycircularvirus Rep and CP sequences, respectively. For smacoviruses, the Blosum62+I+G+F model of substitution was used for Rep sequences and the LG+I+G+F model was used for CP sequences. Branches with less than 80% support were collapsed using TreeGraph 2 (Stöver & Müller, 2010).

Both the Rep and CP sequences of CRESS DNA viruses that do not fall within major groups were analysed using BLASTp against a viral protein database. The CP of a CRESS DNA virus from *R. rupicapra* faeces was found to be most closely related to those encoded by a few RNA-DNA hybrid viruses (Dayaram *et al.*, 2016; Diemer & Stedman, 2012; Krupovic *et*

al., 2015; Roux *et al.*, 2013). The CPs of RNA-DNA hybrid viruses are similar to that of members of the *Tombusviridae* family. CPs of representative tombusviruses and RNA–DNA hybrid viruses were aligned using PROMALS3D (Pie *et al.*, 2008). The resulting alignment was used to infer a maximum-likelihood phylogenetic tree using PHYML (Guindon *et al.*, 2010) with substitution model Blosum62+I+G+F chosen as the best model using ProtTest (Darriba *et al.*, 2011). Branches with aLRT branch support with less than 80% support were collapsed.

Table 2.2: Description of gemycircularviruses identified as of the 5th of January 2016, including isolates recovered in this study.

Accession #	Gemycircularviruses description	Acronym	Country	Isolation source	Common name	Sample type	Reference
GQ365709	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	China	*Sclerotinia sclerotiorum*	*Sclerotinia sclerotiorum*	Mycelial samples	(Yu *et al.*, 2010)
HQ335086	Mosquito VEM virus SDBVL G	MVemV	USA	*Culex erythrothorax*	Mosquito	Mosquito samples	(Ng *et al.*, 2011b)
JN704610	Meles meles fecal virus	MmFV	Netherlands	*Meles meles*	European badger	Rectal swab	(van den Brand *et al.*, 2012)
JQ412057	Cassava associated circular DNA virus	CasCV	Ghana	*Manihot esculenta*	Cassava	Leaf	(Dayaram *et al.*, 2012)
JX185428	Dragonfly-associated circular virus 3	DfasCV-3	Tonga	*Pantala flavescens*	Dragonfly	Abdomen	(Rosario *et al.*, 2012a)
JX185429	Dragonfly-associated circular virus 2	DfasCV-2	USA	*Erythemis simplicicollis*	Dragonfly	Abdomen	(Rosario *et al.*, 2012a)
JX185430	Dragonfly-associated circular virus 1	DfasCV-1	USA	*Miathyria marcella*	Dragonfly	Abdomen	(Rosario *et al.*, 2012a)
KF268025	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	New Zealand	River Sediments	-	River Sediments	(Kraberger *et al.*, 2013)
KF268026	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	New Zealand	River Sediments	-	River Sediments	(Kraberger *et al.*, 2013)
KF268027	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	New Zealand	River Sediments	-	River Sediments	(Kraberger *et al.*, 2013)
KF268028	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	New Zealand	River Sediments	-	River Sediments	(Kraberger *et al.*, 2013)
KF371630	Faecal-associated gemycircularvirus-12	FaGmV-12	New Zealand	*Struthio camelus*	Ostrich	Faeces	(Sikorski *et al.*, 2013d)
KF371631	Faecal-associated gemycircularvirus-11	FaGmV-11	New Zealand	*Oryctolagus cuniculus*	Rabbit	Faeces	(Sikorski *et al.*, 2013d)
KF371632	Faecal-associated gemycircularvirus-10	FaGmV-10	New Zealand	*Sturnus vulgaris*	European starling	Faeces	(Sikorski *et al.*, 2013d)
KF371633	Faecal-associated gemycircularvirus-9	FaGmV-9	New Zealand	*Turdus merula*	Blackbird	Faeces	(Sikorski *et al.*, 2013d)
KF371634	Faecal-associated gemycircularvirus-8	FaGmV-8	New Zealand	*Petroica traversi*	Chatham Island black robin	Faeces	(Sikorski *et al.*, 2013d)
KF371635	Faecal-associated gemycircularvirus-7	FaGmV-7	New Zealand	*Anas platyrhynchos*	Mallard duck	Faeces	(Sikorski *et al.*, 2013d)
KF371636	Faecal-associated gemycircularvirus-6	FaGmV-6	New Zealand	*Gerygone albofrontata*	Chatham Island warbler	Faeces	(Sikorski *et al.*, 2013d)
KF371637	Faecal-associated gemycircularvirus-5	FaGmV-5	New Zealand	*Gerygone albofrontata*	Chatham Island warbler	Faeces	(Sikorski *et al.*, 2013d)
KF371638	Faecal-associated gemycircularvirus-4	FaGmV-4	New Zealand	*Arctocephalus forsteri*	New Zealand fur seal	Faeces	(Sikorski *et al.*, 2013d)
KF371639	Faecal-associated gemycircularvirus-3	FaGmV-3	New Zealand	*Gerygone albofrontata*	Chatham Island warbler	Faeces	(Sikorski *et al.*, 2013d)
KF371640	Faecal-associated gemycircularvirus-2	FaGmV-2	New Zealand	*Sus scrofa*	Domestic pig	Faeces	(Sikorski *et al.*, 2013d)
KF371641	Faecal-associated gemycircularvirus-1c	FaGmV-1c	New Zealand	*Turdus merula*	Blackbird	Faeces	(Sikorski *et al.*, 2013d)
KF371642	Faecal-associated gemycircularvirus-1b	FaGmV-1b	New Zealand	*Turdus merula*	Blackbird	Faeces	(Sikorski *et al.*, 2013d)
KF371643	Faecal-associated gemycircularvirus-1a	FaGmV-1a	New Zealand	*Ovis aries*	Sheep	Faeces	(Sikorski *et al.*, 2013d)
KF413620	Hypericum japonicum associated circular DNA virus	HJasCV	Vietnam	*Hypericum japonicum*	Hypericum	Leaf	(Du *et al.*, 2014)
KJ413144	Human genital-associated circular DNA virus-1	HuGaGmC349	South Africa	*Homo sapiens*	Human	Cervical sample	Unpublished
KJ547634	Sewage-associated gemycircularvirus-4	SaGmV-4	New Zealand	Sewage oxidation pond	-	Sewage	(Kraberger *et al.*, 2015a)
KJ547635	Sewage-associated gemycircularvirus-5	SaGmV-5	New Zealand	Sewage oxidation pond	-	Sewage	(Kraberger *et al.*, 2015a)
KJ547636	Sewage-associated gemycircularvirus-6	SaGmV-6	New Zealand	Sewage oxidation pond	-	Sewage	(Kraberger *et al.*, 2015a)
KJ547637	Sewage-associated gemycircularvirus-7a	SaGmV-7a	New Zealand	Sewage oxidation pond	-	Sewage	(Kraberger *et al.*, 2015a)
KJ547638	Sewage-associated gemycircularvirus-8	SaGmV-8	New Zealand	Sewage oxidation pond	-	Sewage	(Kraberger *et al.*, 2015a)
KJ547639	Sewage-associated gemycircularvirus-9	SaGmV-9	New Zealand	Sewage oxidation pond	-	Sewage	(Kraberger *et al.*, 2015a)
KJ547640	Sewage-associated gemycircularvirus-7b	SaGmV-7b	New Zealand	Sewage oxidation pond	-	Sewage	(Kraberger *et al.*, 2015a)
KJ547641	Sewage-associated gemycircularvirus-11	SaGmV-11	New Zealand	Sewage oxidation pond	-	Sewage	(Kraberger *et al.*, 2015a)
KJ547642	Sewage-associated gemycircularvirus-2	SaGmV-2	New Zealand	Sewage oxidation pond	-	Sewage	(Kraberger *et al.*, 2015a)
KJ547643	Sewage-associated gemycircularvirus-3	SaGmV-3	New Zealand	Sewage oxidation pond	-	Sewage	(Kraberger *et al.*, 2015a)
KJ547644	Sewage-associated gemycircularvirus-10a	SaGmV-10a	New Zealand	Sewage oxidation pond	-	Sewage	(Kraberger *et al.*, 2015a)
KJ547645	Sewage-associated gemycircularvirus-10b	SaGmV-10b	New Zealand	Sewage oxidation pond	-	Sewage	(Kraberger *et al.*, 2015a)
KJ641719	Bat gemycircularvirus 23 GD2012	BtMf-CV-23 GD2012	China	*Miniopterus fuliginosus*	Bat	Pharyngeal & rectal swabs	(Wu *et al.*, 2015)
KJ641726	Bat gemycircularvirus 8 NM2013	BtRf-CV-8 NM2013	China	*Rhinolophus ferrumequinum*	Bat	Pharyngeal & rectal swabs	(Wu *et al.*, 2015)
KJ641737	Bat gemycircularvirus Tibet2013	BtRh-CV-6 Tibet2013	China	*Rhinolophus hipposideros*	Bat	Pharyngeal & rectal swabs	(Wu *et al.*, 2015)
KJ938717	Caribou feces-associated gemycircularvirus	FaGmV-13	Canada	*Rangifer tarandus*	Caribou	Faeces	(Ng *et al.*, 2014)
KM510192	Bromus-associated circular DNA virus 3	BasCV-3	New Zealand	*Bromus hordeaceus*	Soft brome / Bull grass	Leaf	(Kraberger *et al.*, 2015b)
KM598382	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	USA	*Ischnura ramburii*	Damselfly	Abdomen	(Dayaram *et al.*, 2015b)
KM598383	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	USA	*Erythemis simplicicollis*	Dragonfly	Abdomen	(Dayaram *et al.*, 2015b)
KM598384	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	USA	*Pantala hymenaea*	Dragonfly	Abdomen	(Dayaram *et al.*, 2015b)

KM598385	Odonata associated gemycircularvirus-1	OdaGmV-1	USA	*Ischnura posita*	Damselfly	Abdomen	(Dayaram *et al.*, 2015b)
KM598386	Odonata associated gemycircularvirus-1	OdaGmV-1	USA	*Pantala hymenaea*	Dragonfly	Abdomen	(Dayaram *et al.*, 2015b)
KM598387	Odonata associated gemycircularvirus-2	OdaGmV-2	USA	*Aeshna multicolor*	Dragonfly	Abdomen	(Dayaram *et al.*, 2015b)
KM598388	Odonata associated gemycircularvirus-2	OdaGmV-2	USA	*Libellula saturata*	Dragonfly	Abdomen	(Dayaram *et al.*, 2015b)
KM821747	Sewage-associated gemycircularvirus-1	SaGmV-1	New Zealand	Sewage oxidation pond	-	Sewage	(Kraberger *et al.*, 2015a)
KP133075	Gemycircularvirus SL1	GemyCV-SL1	Sri Lanka	*Homo sapiens*	Human	Cerebrospinal fluid	(Phan *et al.*, 2015)
KP133076	Gemycircularvirus SL2	GemyCV-SL2	Sri Lanka	*Homo sapiens*	Human	Cerebrospinal fluid	(Phan *et al.*, 2015)
KP133077	Gemycircularvirus SL3	GemyCV-SL3	Sri Lanka	*Homo sapiens*	Human	Cerebrospinal fluid	(Phan *et al.*, 2015)
KP133078	Gemycircularvirus BZ1	GemyCV-BZ1	Brazil	*Homo sapiens*	Human	Faeces	(Phan *et al.*, 2015)
KP133079	Gemycircularvirus BZ2	GemyCV-BZ2	Brazil	*Homo sapiens*	Human	Faeces	(Phan *et al.*, 2015)
KP133080	Gemycircularvirus NP	GemyCV-NP	Nepal	Untreated sewage	-	Sewage	(Phan *et al.*, 2015)
KP263543	Badger faeces-associated gemycircularvirus	BafaGM588	Portugal	*Meles meles*	European badger	Faeces	(Conceicao-Neto *et al.*, 2015)
KP263544	Mongoose feces-associated gemycircularvirus a	MoFaGmV181a	Portugal	*Herpestes ichneumon*	Egyptian mongoose	Faeces	(Conceicao-Neto *et al.*, 2015)
KP263545	Mongoose feces-associated gemycircularvirus b	MoFaGmV160b	Portugal	*Herpestes ichneumon*	Egyptian mongoose	Faeces	(Conceicao-Neto *et al.*, 2015)
KP263546	Mongoose feces-associated gemycircularvirus c	MoFaGmV541c	Portugal	*Herpestes ichneumon*	Egyptian mongoose	Faeces	(Conceicao-Neto *et al.*, 2015)
KP263547	Mongoose feces-associated gemycircularvirus d	MoFaGmV478d	Portugal	*Herpestes ichneumon*	Egyptian mongoose	Faeces	(Conceicao-Neto *et al.*, 2015)
KP987887	Gemycircularvirus C1c	C1c	France	*Homo sapiens*	Human	Plasma	(Uch *et al.*, 2015)
KR912221	Gemycircularvirus gemy-ch-rat1	Gemy-ch-rat1	China	*Rattus norvegicus*	Rat	Blood	(Li *et al.*, 2015)
KT253577	Poaceae associated gemycircularvirus-1	PaGmV-1	Tonga	*Brachiaria deflexa*	Signalgrass	Leaf	(Male *et al.*, 2015)
KT253578	Poaceae associated gemycircularvirus-1	PaGmV-1	Tonga	*Brachiaria deflexa*	Signalgrass	Leaf	(Male *et al.*, 2015)
KT253579	Poaceae associated gemycircularvirus-1	PaGmV-1	Tonga	*Saccharum hybrid*	Sugarcane	Leaf	(Male *et al.*, 2015)
KT309029	Poecile atricapillus GI tract-associated gemycircularvirus	Gitract	USA	*Poecile atricapillus*	Black-capped chickadee	Buccal and cloacal swab	(Hanna *et al.*, 2015)
KT598248	Soybean leaf-associated gemycircularvirus 1	SlaGemV1	USA	*Glycine max*	Soybean	Leaf	(Marzano & Domier, 2015)
KT732790	Pacific flying fox faeces associated gemycircularvirus-1	PfffaGmV-1	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732791	Pacific flying fox faeces associated gemycircularvirus-1	PfffaGmV-1	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732792	Pacific flying fox faeces associated gemycircularvirus-2	PfffaGmV-2	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732793	Pacific flying fox faeces associated gemycircularvirus-2	PfffaGmV-2	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732794	Pacific flying fox faeces associated gemycircularvirus-3	PfffaGmV-3	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732795	Pacific flying fox faeces associated gemycircularvirus-4	PfffaGmV-4	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732796	Pacific flying fox faeces associated gemycircularvirus-4	PfffaGmV-4	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732797	Pacific flying fox faeces associated gemycircularvirus-5	PfffaGmV-5	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732798	Pacific flying fox faeces associated gemycircularvirus-6	PfffaGmV-6	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732799	Pacific flying fox faeces associated gemycircularvirus-6	PfffaGmV-6	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732800	Pacific flying fox faeces associated gemycircularvirus-7	PfffaGmV-7	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732801	Pacific flying fox faeces associated gemycircularvirus-8	PfffaGmV-8	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732802	Pacific flying fox faeces associated gemycircularvirus-8	PfffaGmV-8	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732803	Pacific flying fox faeces associated gemycircularvirus-9	PfffaGmV-9	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732804	Pacific flying fox faeces associated gemycircularvirus-10	PfffaGmV-10	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732805	Pacific flying fox faeces associated gemycircularvirus-10	PfffaGmV-10	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732807	Pacific flying fox faeces associated gemycircularvirus-11	PfffaGmV-11	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732808	Pacific flying fox faeces associated gemycircularvirus-11	PfffaGmV-11	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732809	Pacific flying fox faeces associated gemycircularvirus-11	PfffaGmV-11	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732810	Pacific flying fox faeces associated gemycircularvirus-11	PfffaGmV-11	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732811	Pacific flying fox faeces associated gemycircularvirus-11	PfffaGmV-11	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732812	Pacific flying fox faeces associated gemycircularvirus-11	PfffaGmV-11	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732813	Pacific flying fox faeces associated gemycircularvirus-12	PfffaGmV-12	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732814	Pacific flying fox faeces associated gemycircularvirus-13	PfffaGmV-13	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT732806	Pacific flying fox faeces associated gemycircularvirus-14	PfffaGmV-14	Tonga	*Pteropus tonganus*	Bat	Faeces	(Male *et al.*, 2016)
KT862238	Faecal-associated gemycircularvirus-14	FaGmV-14	New Zealand	*Anas platyrhynchos*	Duck	Faeces	This study
KT862239	Faecal-associated gemycircularvirus-14	FaGmV-14	New Zealand	*Anas platyrhynchos*	Duck	Faeces	This study
KT862240	Sewage-associated gemycircularvirus-3	SaGmV-3	New Zealand	*Gallus gallus domesticus*	Chicken	Faeces	This study
KT862241	Faecal-associated gemycircularvirus-4	FaGmV-4	New Zealand	*Gallus gallus domesticus*	Chicken	Faeces	This study
KT862242	Faecal-associated gemycircularvirus-17	FaGmV-17	New Zealand	*Gallus gallus domesticus*	Chicken	Faeces	This study

KT862243	Faecal-associated gemycircularvirus-20	FaGmV-20	New Zealand	*Gallus gallus domesticus*	Chicken	Faeces	This study
KT862244	Faecal-associated gemycircularvirus-20	FaGmV-20	New Zealand	*Lama glama*	Llama	Faeces	This study
KT862245	Faecal-associated gemycircularvirus-21	FaGmV-21	New Zealand	*Lama glama*	Llama	Faeces	This study
KT862246	Faecal-associated gemycircularvirus-20	FaGmV-20	New Zealand	*Equus ferus caballus*	Horse	Faeces	This study
KT862247	Faecal-associated gemycircularvirus-21	FaGmV-21	New Zealand	*Equus ferus caballus*	Horse	Faeces	This study
KT862248	Faecal-associated gemycircularvirus-18	FaGmV-18	New Zealand	*Equus ferus caballus*	Horse	Faeces	This study
KT862249	Faecal-associated gemycircularvirus-16	FaGmV-16	New Zealand	*Ovis aries*	Sheep	Faeces	This study
KT862250	Faecal-associated gemycircularvirus-19	FaGmV-19	New Zealand	*Sus scrofa domesticus*	Pig	Faeces	This study
KT862251	Faecal-associated gemycircularvirus-16	FaGmV-16	New Zealand	*Ovis aries*	Sheep	Faeces	This study
KT862252	Sewage-associated gemycircularvirus-3	SaGmV-3	New Zealand	*Bos taurus*	Cow	Faeces	This study
KT862253	Faecal-associated gemycircularvirus-22	FaGmV-22	New Zealand	*Bos taurus*	Cow	Faeces	This study
KT862254	Faecal-associated gemycircularvirus-15	FaGmV-15	New Zealand	*Canis lupus familiaris*	Dog	Faeces	This study
KT862255	Sewage-associated gemycircularvirus-3	SaGmV-3	New Zealand	*Lepus europaeus*	Hare	Faeces	This study
LK931483	HCBI8.215 virus	HCBI8_215	Germany	*Bos taurus*	Cow	Serum	(Lamberto *et al.*, 2014)
LK931484	HCBI9.212 virus	HCBI9_212	Germany	*Bos taurus*	Cow	Serum	(Lamberto *et al.*, 2014)
LK931485	MSSI2.225 virus	MSSI2_225	Germany	*Homo sapiens*	Human	Blood	(Lamberto *et al.*, 2014)

Table 2.3: Description of smacoviruses identified as of the 5th of January 2016, including isolates recovered in this study.

Accession #	Smacovirus description	Acronym	Clade grouping	Genome type	Country	Faecal source	Common name	Reference
GQ351272	Chimpanzee stool associated circular ssDNA virus	ChiSCV	IX	Bidirectional	Cameroon	*Pan troglodytes*	Chimpanzee	(Blinkova *et al.*, 2010)
GQ351273	Chimpanzee stool associated circular ssDNA virus	ChiSCV	IX	Bidirectional	Tanzania	*Pan troglodytes*	Chimpanzee	(Blinkova *et al.*, 2010)
GQ351274	Chimpanzee stool associated circular ssDNA virus	ChiSCV	IX	Bidirectional	Tanzania	*Pan troglodytes*	Chimpanzee	(Blinkova *et al.*, 2010)
GQ351275	Chimpanzee stool associated circular ssDNA virus	ChiSCV	IX	Bidirectional	Tanzania	*Pan troglodytes*	Chimpanzee	(Blinkova *et al.*, 2010)
GQ351276	Chimpanzee stool associated circular ssDNA virus	ChiSCV	IX	Bidirectional	Tanzania	*Pan troglodytes*	Chimpanzee	(Blinkova *et al.*, 2010)
GQ351277	Chimpanzee stool associated circular ssDNA virus	ChiSCV	IX	Bidirectional	Tanzania	*Pan troglodytes*	Chimpanzee	(Blinkova *et al.*, 2010)
JN634851	Bovine stool-associated circular virus	BoSCV	III	Bidirectional	South Korea	*Bos taurus*	Cow	(Kim *et al.*, 2012)
JQ023166	Pig stool associated circular ssDNA virus	PigSCV	II	Unidirectional	Germany	*Sus scrofa domesticus*	Pig	(Sachsenröder *et al.*, 2012)
JX274036	Porcine stool-associated circular virus	PoSCV	XI	Bidirectional	New Zealand	*Sus scrofa domesticus*	Pig	(Sikorski *et al.*, 2013a)
JX305991	Pig stool associated circular ssDNA virus	PigSCV	II	Unidirectional	China	*Sus scrofa domesticus*	Pig	Unpublished
JX305992	Pig stool associated circular ssDNA virus	PigSCV	II	Unidirectional	China	*Sus scrofa domesticus*	Pig	Unpublished
JX305993	Pig stool associated circular ssDNA virus	PigSCV	II	Unidirectional	China	*Sus scrofa domesticus*	Pig	Unpublished
JX305994	Pig stool associated circular ssDNA virus	PigSCV	II	Unidirectional	China	*Sus scrofa domesticus*	Pig	Unpublished
JX305995	Pig stool associated circular ssDNA virus	PigSCV	II	Unidirectional	China	*Sus scrofa domesticus*	Pig	Unpublished
JX305996	Pig stool associated circular ssDNA virus	PigSCV	II	Unidirectional	China	*Sus scrofa domesticus*	Pig	Unpublished
JX305997	Pig stool associated circular ssDNA virus	PigSCV	II	Unidirectional	China	*Sus scrofa domesticus*	Pig	Unpublished
JX305998	Pig stool associated circular ssDNA virus	PigSCV	II	Unidirectional	China	*Sus scrofa domesticus*	Pig	Unpublished
KC545226	Porcine stool-associated circular virus-2	PoSCV-2	VI	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2013)
KC545227	Porcine stool-associated circular virus-3	PoSCV-3	VI	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2013)
KC545228	Porcine stool-associated circular virus-3	PoSCV-3	VI	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2013)
KC545229	Porcine stool-associated circular virus-3	PoSCV-3	VI	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2013)
KC545230	Porcine stool-associated circular virus-3	PoSCV-3	VI	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2013)
KF193403	Pig stool associated circular ssDNA virus	PoSCV	XI	Bidirectional	South Korea	*Sus scrofa domesticus*	Pig	(Kim *et al.*, 2014)
KF880727	Turkey stool associated circular ssDNA virus	TuSCV	XI	Bidirectional	Hungary	*Meleagris gallopavo*	Turkey	(Reuter *et al.*, 2014)
KJ547633	Sewage-associated circular virus-9	SaCV-9	II	Unidirectional	New Zealand	Sewage	Sewage	(Kraberger *et al.*, 2015a)
KJ577810	Porcine stool-associated circular virus-1	PoSCV-1	XI	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577811	Porcine stool-associated circular virus-1	PoSCV-1	XI	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577812	Porcine stool-associated circular virus-7	PoSCV-7	VII	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577813	Porcine stool-associated circular virus-7	PoSCV-7	VII	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577814	Porcine stool-associated circular virus-7	PoSCV-7	VII	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577815	Porcine stool-associated circular virus-7	PoSCV-7	VII	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577816	Porcine stool-associated circular virus-9	PoSCV-9	IV	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577817	Porcine stool-associated circular virus-8	PoSCV-8	V	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577818	Porcine stool-associated circular virus-2	PoSCV-2	VI	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577819	Porcine stool-associated circular virus-6	PoSCV-6	VIII	Bidirectional	USA	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KM573769	Dromedary stool-associated circular ssDNA virus	DcSCV	XII	Bidirectional	UAE	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573770	Dromedary stool-associated circular ssDNA virus	DcSCV	XXI	Bidirectional	UAE	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573771	Dromedary stool-associated circular ssDNA virus	DcSCV	XVIII	Bidirectional	UAE	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573772	Dromedary stool-associated circular ssDNA virus	DcSCV	XXII	Bidirectional	UAE	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573774	Dromedary stool-associated circular ssDNA virus	DcSCV	XII	Bidirectional	UAE	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573775	Dromedary stool-associated circular ssDNA virus	DcSCV	XVIII	Bidirectional	UAE	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM598409	Odonata-associated circular virus-21	OdasCV-21	XVI	Bidirectional	USA	*Erythrodiplax fusca*	Dragonfly	(Dayaram *et al.*, 2015b)
KP233174	Human smacovirus-1	HuSCV-1	I	Bidirectional	France	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233175	Human smacovirus-1	HuSCV-1	I	Bidirectional	France	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233176	Human smacovirus-1	HuSCV-1	I	Bidirectional	France	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233177	Human smacovirus-1	HuSCV-1	I	Bidirectional	France	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233178	Human smacovirus-1	HuSCV-1	I	Bidirectional	France	*Homo sapiens*	Human	(Ng *et al.*, 2015)

KP233179	Human smacovirus-1	HuSCV-1	I	Bidirectional	France	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233180	Human smacovirus-1	HuSCV-1	I	Bidirectional	USA	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233181	Human smacovirus-1	HuSCV-1	I	Bidirectional	USA	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233182	Human smacovirus-1	HuSCV-1	I	Bidirectional	USA	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233183	Human smacovirus-1	HuSCV-1	I	Bidirectional	USA	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233184	Human smacovirus-1	HuSCV-1	I	Bidirectional	USA	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233185	Human smacovirus-1	HuSCV-1	I	Bidirectional	USA	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233186	Human smacovirus-1	HuSCV-1	I	Bidirectional	USA	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233187	Human smacovirus-1	HuSCV-1	I	Bidirectional	USA	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233188	Human smacovirus-1	HuSCV-1	I	Bidirectional	USA	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233189	Black howler monkey smacovirus	BhmSCV	XI	Bidirectional	USA	*Alouatta caraya*	Black howler monkey	(Ng *et al.*, 2015)
KP233190	Chimpanzee stool associated circular ssDNA virus	ChiSCV	XI	Bidirectional	USA	*Pan troglodytes*	Chimpanzee	(Ng *et al.*, 2015)
KP233191	Gorilla smacovirus	GoSCV	VII	Bidirectional	USA	*Gorilla gorilla*	Gorilla	(Ng *et al.*, 2015)
KP233192	Gorilla smacovirus	GoSCV	VII	Bidirectional	USA	*Gorilla gorilla*	Gorilla	(Ng *et al.*, 2015)
KP233193	Human smacovirus-1	HuSCV-1	I	Bidirectional	USA	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233194	Lemur smacovirus	LeSCV	X	Bidirectional	USA	*Lemur catta*	Lemur	(Ng *et al.*, 2015)
KP264964	Human smacovirus-1	HuSCV-1	I	Bidirectional	France	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP264965	Human smacovirus-1	HuSCV-1	I	Bidirectional	France	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP264966	Human smacovirus-1	HuSCV-1	I	Bidirectional	France	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP264967	Human smacovirus-1	HuSCV-1	I	Bidirectional	France	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP264968	Human smacovirus-1	HuSCV-1	I	Bidirectional	France	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP264969	Human smacovirus-1	HuSCV-1	I	Bidirectional	France	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP860906	Rat stool-associated circular ssDNA virus	RsaCV	VIII	Bidirectional	Germany	*Rattus norvegicus* *	Rat	(Sachsenröder *et al.*, 2014)
KP860907	Rat stool-associated circular ssDNA virus	RsaCV	VIII	Bidirectional	Germany	*Rattus norvegicus* *	Rat	(Sachsenröder *et al.*, 2014)
KP860908	Rat stool-associated circular ssDNA virus	RsaCV	VIII	Bidirectional	Germany	*Rattus norvegicus* *	Rat	(Sachsenröder *et al.*, 2014)
KT862218	Bovine faeces associated smacovirus -2	BofSmV-2	XXII	Bidirectional	New Zealand	*Bos taurus*	Cow	This study
KT862219	Sheep faeces associated smacovirus -3	ShfSmV-3	XVII	Bidirectional	New Zealand	*Ovis aries*	Sheep	This study
KT862220	Sheep faeces associated smacovirus -1	ShfSmV-1	XIX	Bidirectional	New Zealand	*Ovis aries*	Sheep	This study
KT862221	Sheep faeces associated smacovirus -2	ShfSmV-2	XXIII	Bidirectional	New Zealand	*Ovis aries*	Sheep	This study
KT862222	Bovine faeces associated smacovirus -3	BofSmV-3	III	Bidirectional	New Zealand	*Bos taurus*	Cow	This study
KT862223	Bovine faeces associated smacovirus -1	BofSmV-1	XIII	Bidirectional	New Zealand	*Bos taurus*	Cow	This study
KT862224	Bovine faeces associated smacovirus -5	BofSmV-5	XII	Bidirectional	New Zealand	*Bos taurus*	Cow	This study
KT862225	Porcine faeces associated smacovirus -1	PofSmV-1	XX	Bidirectional	New Zealand	*Sus scrofa domesticus*	Pig	This study
KT862226	Porcine stool-associated circular virus	PoSCV	XI	Bidirectional	New Zealand	*Lepus europaeus*	Hare	This study
KT862227	Porcine stool-associated circular virus	PoSCV	XI	Bidirectional	New Zealand	*Trichosurus vulpecula*	Common brushtail possum	This study
KT862228	Bovine faeces associated smacovirus -4	BofSmV-4	XV	Bidirectional	New Zealand	*Bos taurus*	Cow	This study
KT862229	Bovine faeces associated smacovirus -6	BofSmV-6	XIV	Bidirectional	New Zealand	*Bos taurus*	Cow	This study

2.4 Results and discussion

2.4.1 Identification and classification of CRESS DNA viruses

A viral metagenomics study of animal faecal samples (n=49) collected from both wild and domestic animals between 2009 and 2013 was conducted to explore the diversity of CRESS DNA viruses on the South Island of New Zealand. Through analysis of *de novo* assembled contigs of the Illumina sequencing paired-end reads, 708 contigs were identified that were larger than 500nts and had viral-like hits in GenBank's non-redundant database. Of these, 31% were found to have to CRESS DNA virus-like sequences.

Abutting primers were designed to recover putative viral circular DNA molecules from the animal faecal samples (Table 2.1). As part of this study, 40 CRESS DNA molecules (Figure 2.1; Table 2.4) were recovered from nineteen faecal samples (*R. rupicapra*, n=1; *G. gallus domesticus* n=1; *T. vulpecula*, n=1; *B. taurus*, n=4; *D. dama*, n=1; *C. lupus familiaris*, n=1; *A. platyrhynchos*, n=3; *L. europaeus*, n=1; *E. ferus*, n=1; *L. glama*, n=1; *S. scrofa domesticus*, n=2; *O. aries*, n=2). From seven samples only one CRESS DNA virus was identified, however, in the remaining samples two - four CRESS DNA viruses were identified (Figure 2.1; Table 2.4).

Two of the molecules identified are small (1280 and 1274nts) and have one large ORF that encodes the Rep. The remaining 38 circular molecules (1845 – 3060nts) encode at least two large ORFs, both with similarities to Reps and CPs encoded by other CRESS DNA viruses. These 38 CRESS DNA viruses have been grouped into gemycircularviruses (n=18), smacoviruses (n=12) and unclassified CRESS DNA viruses (n=8) (Figure 2.1; Table 2.4).

The genome organisation of CRESS DNA viruses differs with respect to ORF orientation, position of the stem-loop relative to the ORFs and orientation of the origin of replication relative to the Rep ORF (Rosario *et al.*, 2012b). A classification scheme described by Rosario *et al.* (2012b) classifies CRESS DNA virus genomes into one of eight groups based on these characteristics (Table 1.2). All gemycircularviruses identified in this study have a type II genome architecture, while the smacoviruses have a type IV genome organisation

Figure 2.1: Summary of CRESS DNA viruses and circular molecules recovered in the present study from wild and domestic animal faecal sources collected from across the South Island of New Zealand

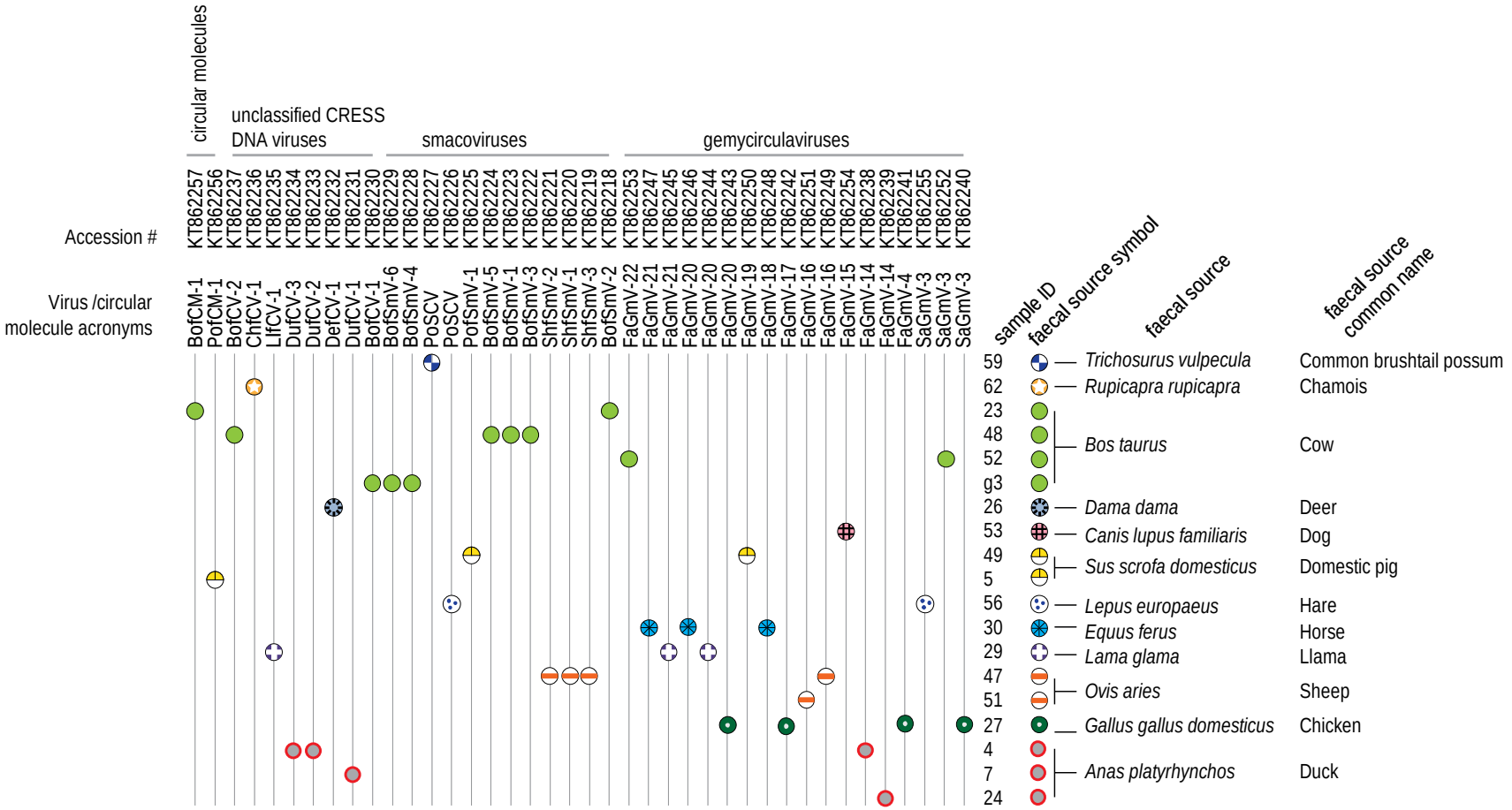


Table 2.4: Grouping of CRESS DNA viruses recovered in this study as gemycircularviruses, smacoviruses, unclassified CRESS DNA viruses and circular molecules.

Grouping	Accession #	Virus / circular molecule	Acronym	Faecal source	Faecal source common name	Sampling year
Gemycircularviruses	KT862238	Faeces associated gemycircularvirus-14	FaGmV-14	*Anas platyrhynchos*	Duck	2012
	KT862239	Faeces associated gemycircularvirus-14	FaGmV-14	*Anas platyrhynchos*	Duck	2012
	KT862240	Sewage associated gemycircularvirus-3	SaGmV-3	*Gallus gallus domesticus*	Chicken	2012
	KT862241	Faeces associated gemycircularvirus-4	FaGmV-4	*Gallus gallus domesticus*	Chicken	2012
	KT862242	Faeces associated gemycircularvirus-17	FaGmV-17	*Gallus gallus domesticus*	Chicken	2012
	KT862243	Faeces associated gemycircularvirus-20	FaGmV-20	*Gallus gallus domesticus*	Chicken	2012
	KT862244	Faeces associated gemycircularvirus-20	FaGmV-20	*Lama glama*	Llama	2012
	KT862245	Faeces associated gemycircularvirus-21	FaGmV-21	*Lama glama*	Llama	2012
	KT862246	Faeces associated gemycircularvirus-20	FaGmV-20	*Equus ferus*	Horse	2012
	KT862247	Faeces associated gemycircularvirus-21	FaGmV-21	*Equus ferus*	Horse	2012
	KT862248	Faeces associated gemycircularvirus-18	FaGmV-18	*Equus ferus*	Horse	2012
	KT862249	Faeces associated gemycircularvirus-16	FaGmV-16	*Ovis aries*	Sheep	2012
	KT862250	Faeces associated gemycircularvirus-19	FaGmV-19	*Sus scrofa domesticus*	Pig	2012
	KT862251	Faeces associated gemycircularvirus-16	FaGmV-16	*Ovis aries*	Sheep	2012
	KT862252	Sewage associated gemycircularvirus-3	SaGmV-3	*Bos taurus*	Cow	2012
	KT862253	Faeces associated gemycircularvirus-22	FaGmV-22	*Bos taurus*	Cow	2012
	KT862254	Faeces associated gemycircularvirus-15	FaGmV-15	*Canis lupus familiaris*	Dog	2012
	KT862255	Sewage associated gemycircularvirus-3	SaGmV-3	*Lepus europaeus*	Hare	2011
Smacoviruses	KT862218	Bovine faeces associated smacovirus -2	BofSmV-2	*Bos taurus*	Cow	2012
	KT862219	Sheep faeces associated smacovirus -3	ShfSmV-3	*Ovis aries*	Sheep	2012
	KT862220	Sheep faeces associated smacovirus -1	ShfSmV-1	*Ovis aries*	Sheep	2012
	KT862221	Sheep faeces associated smacovirus -2	ShfSmV-2	*Ovis aries*	Sheep	2012
	KT862222	Bovine faeces associated smacovirus -3	BofSmV-3	*Bos taurus*	Cow	2012
	KT862223	Bovine faeces associated smacovirus -1	BofSmV-1	*Bos taurus*	Cow	2012
	KT862224	Bovine faeces associated smacovirus -5	BofSmV-5	*Bos taurus*	Cow	2012
	KT862225	Porcine faeces associated smacovirus -1	PofSmV-1	*Sus scrofa domesticus*	Pig	2012
	KT862226	Porcine stool-associated circular virus	PoSCV	*Lepus europaeus*	Hare	2011
	KT862227	Porcine stool-associated circular virus	PoSCV	*Trichosurus vulpecula*	Common brushtail possum	2011
	KT862228	Bovine faeces associated smacovirus -4	BofSmV-4	*Bos taurus*	Cow	2013
	KT862229	Bovine faeces associated smacovirus -6	BofSmV-6	*Bos taurus*	Cow	2013
Unclassified CRESS DNA viruses	KT862230	Bovine faeces associated circular DNA virus-1	BofCV-1	*Bos taurus*	Cow	2013
	KT862231	Duck faeces associated circular DNA virus-1	DufCV-1	*Anas platyrhynchos*	Duck	2012
	KT862232	Deer faeces associated circular DNA virus-1	DefCV-1	*Dama dama*	Deer	2012
	KT862233	Duck faeces associated circular DNA virus-2	DufCV-2	*Anas platyrhynchos*	Duck	2011
	KT862234	Duck faeces associated circular DNA virus-3	DufCV-3	*Anas platyrhynchos*	Duck	2011
	KT862235	Llama faeces associated circular DNA virus-1	LlfCV-1	*Lama glama*	Llama	2012
	KT862236	Chamois faeces associated circular DNA virus-1	ChfCV-1	*Rupicapra rupicapra*	Chamois	2011
	KT862237	Bovine faeces associated circular DNA virus-2	BofCV-2	*Bos taurus*	Cow	2012
Circular molecules	KT862256	Porcine faeces associated circular DNA molecule-1	PofCM-1	*Sus scrofa domesticus*	Pig	2011
	KT862257	Bovine faeces associated circular DNA molecule-1	BofCM-1	*Bos taurus*	Cow	2012

(Figure 2.2). The genome architectures of unclassified CRESS DNA viruses are more varied with type I, III, IV, and V genome architectures represented in this study. The two circular molecules identified have a type VII genome architecture.

The 38 CRESS DNA viruses and two circular molecules are discussed in detail within their groupings below.

2.4.2 Gemycircularviruses

Gemycircularviruses have approximately 2.0 - 2.4kb circular ssDNA genomes that encode at least two ORFs and have a type II circular genome organisation (Figure 2.2; Table 1.2) (Rosario *et al.*, 2012b) i.e. the CP is encoded on the virion-sense strand and the Rep is encoded on the complementary-sense strand. The ORFs overlap at their 3' ends, therefore the gemycircularviruses only contain a single intergenic region (IR). Both ORFs are transcribed away from the putative origin of replication located in the IR. The putative origin of replication contains the characteristic stem-loop structure with a nonanucleotide motif positioned on the virion-sense strand. Furthermore, the majority of the gemycircularviruses appear to encode the Rep from a spliced *rep* transcript (Roumagnac *et al.*, 2015; Varsani *et al.*, 2014b; Wright *et al.*, 1997). Some geminiviruses express both an unspliced *rep* transcript (*repA*) and spliced *rep* transcript that are both important for viral proliferation (Dekker *et al.*, 1991; Liu *et al.*, 1998; Wright *et al.*, 1997). The Rep of gemycircularviruses is most closely related to those of geminiviruses. The host of only one gemycircularvirus is known (Yu *et al.*, 2010). Based on this and coupled with the fact that Rep-like sequences have been identified in fungal genomes (Liu *et al.*, 2011), it is highly likely that fungi may be the hosts of gemycircularviruses.

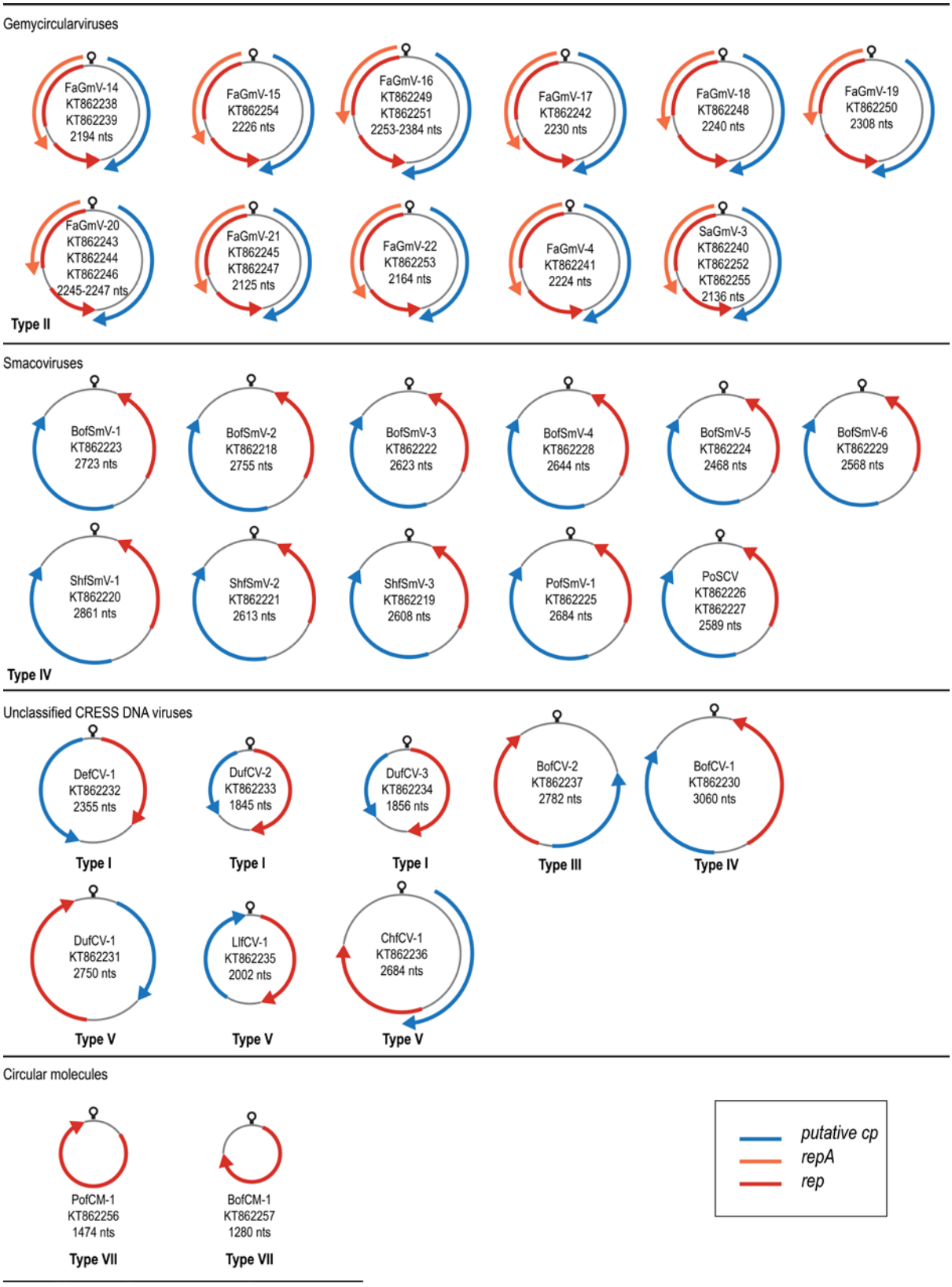
In this study, eighteen gemycircularvirus genomes were recovered from twelve animal faecal samples collected on the South Island of New Zealand (Figure 2.1; Table 2.4). Based on the genome-wide nucleotide pairwise identities, these gemycircularviruses represent eleven putative species, nine of which are novel (Figure 2.1; Table 2.4). An 80% pairwise identity species cut-off was applied, as proposed by Sikorski *et al.* (2013d) and Kraberger *et al.* (2015a). The nine novel gemycircularvirus species have been tentatively named Faeces-

associated gemycircularvirus (FaGmV)-14 (n=2), -15 (n=1), -16 (n=2), -17 (n=1), -18 (n=1), -19 (n=1), -20 (n=3), -21 (n=2) and -22 (n=1), following on the naming by Sikorski *et al.* (2013d) who identified gemycircularviruses in faecal samples of *A. platyrhynchos*, *Arctocephalus forsteri*, *Gerygone albofrontata*, *Oryctolagus cuniculus*, *O. aries*, *Petroica traversi*, *Struthio camelus*, *Stumus vulgaris*, *S. scrofa* and *Turdus merula* (FaGmV-1 to -12), and Ng *et al.* (2014) who identified a gemycircularvirus in *Rangifer tarandus* (FaGmV-13).

Three gemycircularviruses recovered from *B. taurus*, *G. gallus domesticus* and *L. europaeus* share ~94% genome-wide nucleotide pairwise identity with Sewage-associated gemycircularvirus-3 (SaGmV-3, KJ547643), therefore these sequences have been named SaGmV-3 (KT862240, KT862252, KT862255). A gemycircularvirus recovered from *G. gallus domesticus* in this study shares 99.8% identity with FaGmV-4 (KF371638) recovered from *A. forsteri* in New Zealand, so has been named FaGmV-4 (KT862241). Additionally, a Meles meles fecal virus isolate VS4700006 (MmFV, JN704610) previously identified in the Netherlands shares 98.4% genome-wide identity with FaGmV-14 isolates (KT862238 - KT862239) identified in two *A. platyrhynchos* faecal samples. The nonanucleotide motif of the gemycircularviruses from this study varies with TAATATTAT and TAATGTTAT being the most common (Table 2.5). The nonanucleotide motifs of other gemycircularvirus isolates are TATATAAAG, TATAAATAC, TAATGCTAT and TAATACTAT.

The Reps of the eighteen gemycircularviruses identified in this study contain a putative intron (Figure 2.2), similar to other gemycircularviruses except SsHADV-1. These introns are bordered by canonical acceptor (GT) and donor (AG) sites, and range from 130 - 190nts in this study. The size of the spliced *rep* and *repA* transcripts are 982 - 1188nts and 602 - 758nts, respectively. The CP ORFs of recovered isolates are between 906 and 1029nts The Rep proteins of gemycircularviruses from this study have RCR motif I, II and III, which are highly conserved (Table 2.5). RCR motif I was identified as LuTYxQ, with the sequence LLTYAQ detected in 14 of the sequences. xHuHx was identified as RCR motif II and RCR motif III was identified as YAxK. RCR motif III was highly conserved among isolates, with the sequence YAIK detected in 15 isolates. Identification of a single tyrosine residue in RCR motif III classifies the encoded Rep proteins as members of Rep superfamily II. Consistent with previous reports, the GRS identified lacked the QxAK sequence present in the GRS of

Figure 2.2: Genome organisations of gemycircularviruses, smacoviruses, unclassified CRESS DNA viruses and circular molecules recovered in the present study from the faecal matter of wild and domestic animals in New Zealand.



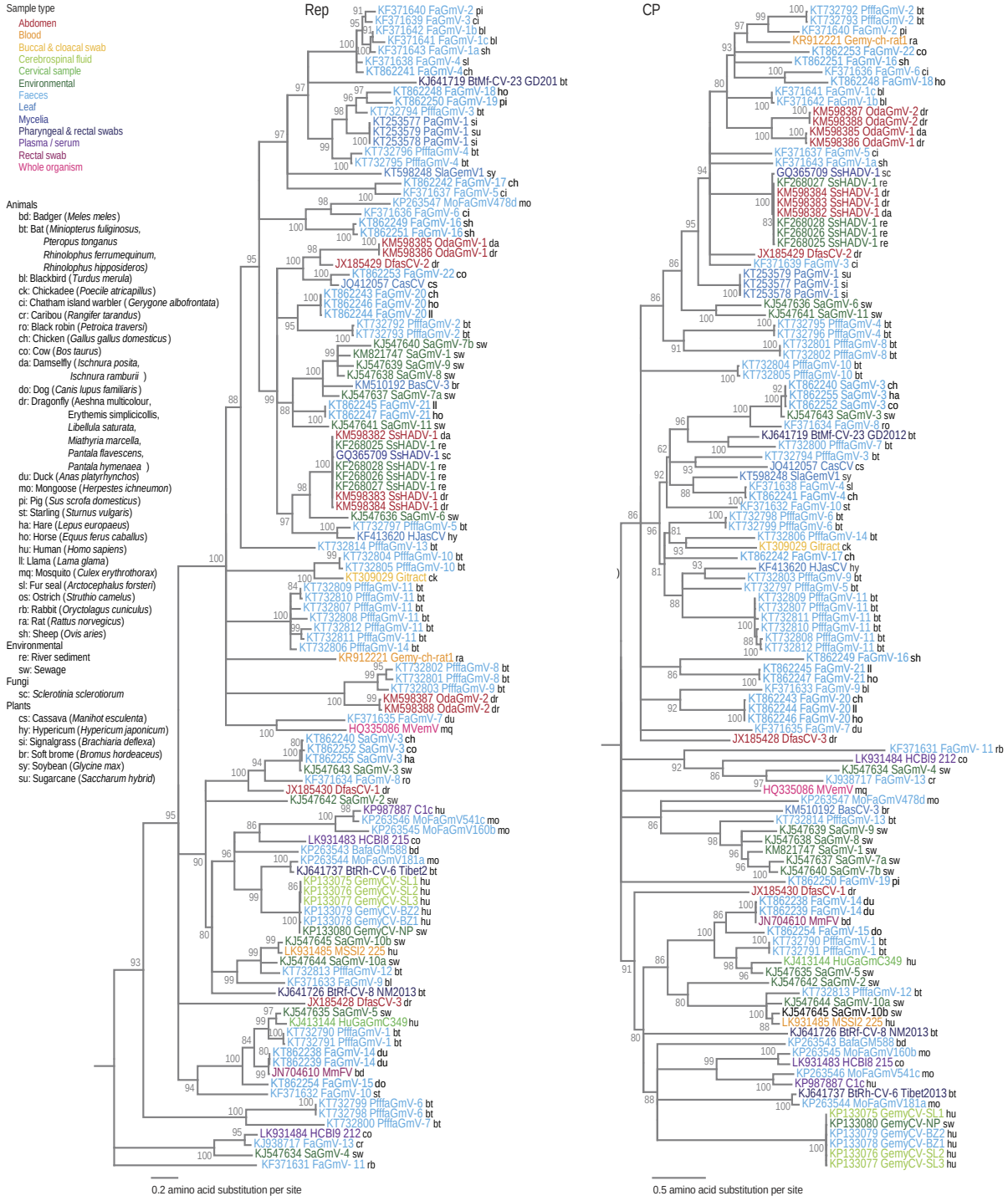
geminiviruses (Rosario *et al.*, 2012b). Walker A, B and motif C were identified in the NTP-binding domain of putative Reps in recovered gemycircularviruses (Table 2.5). In the recovered genomes, the Walker A motif is GxxxxGKT and Walker B is uuxu(D/E)(D/E)u. The most common Walker B motif identified in this study is VFDDM, which is found in FaGmV-4, FaGmV-16 isolates, FaGmV-17, FaGmV-19 and SaGmV-3 isolates. Motif C in the helicase domain is uxxN, most commonly WLAN. The presence of these motifs suggests that the putative Reps have both endonuclease and helicase properties.

Overall the full genomes of gemycircularviruses identified to date have ~45% diversity. The gemycircularviruses have high diversity in their CP (~80%) and Rep (~75%) amino acid sequences. Maximum-likelihood phylogenetic trees constructed using alignments of the putative Rep and CP amino acid sequences are shown in Figure 2.3. In general, the Reps are more conserved that the CPs. The Reps of FaGmV-4 (KF371638, KT862241) are most closely related to those of FaGmV-1, -2 and -3 (KF371639 - KF371643) sharing 81-88% pairwise amino acid identity, however, their CPs are most closely related to soybean leaf-associated gemycircularvirus 1 (SlaGemV1-1; KT598248) sharing ~55% pairwise amino acid identity. The Rep sequences of FaGmV-21 isolates cluster with gemycircularviruses recovered from a sewage oxidation pond sample (SaGmV-1, KM821747; SaGmV-7, KJ547637, KJ547640; SaGmV-8, KJ547638; SaGmV-9, KJ547639) and *Bromus hordeaceus* (BasCV, KM510192) sharing 67-70% amino acid identity, which is not observed in the CP maximum-likelihood phylogenetic tree. The Reps of FaGmV-16 (KT862249, KT862251) share 96% pairwise amino acid identity, however their CPs are distantly related sharing 35% identity.

Table 2.5: Conserved RCR and SF3 helicase motifs in the Reps of gemycircularviruses, smacoviruses, unclassified CRESS DNA viruses and circular molecules recovered in this study.

		Accession #	Nonanucleotide motif	Motif I	Motif II	Motif III	GRS	Walker A	Walker B	Motif C
Gemycircularviruses	FaGmV-14	KT862238	TATATAAAG	LLTYAQ	IHLHA	YAIK	RRFDVGGFHPNIAPCG	GETRLGKT	VLDDM	WLMN
	FaGmV-14	KT862239	TATATAAAG	LLTYAQ	IHLHA	YAIK	RRFDVGGFHPNIAPCG	GETRLGKT	VLDDM	WLMN
	SaGmV-3	KT862240	TAATGTTAT	LLTYAQ	LHIHA	YAIK	RVFDMDGCHPNIVRGY	GPTKLGKT	VFDDM	YLYN
	FaGmV-4	KT862241	TAATGTTAT	LLTYAQ	THLHA	YAIK	DVFDVGGFHPNIEASR	GDTRLGKT	VFDDM	WLAN
	FaGmV-17	KT862242	TAATATTAT	LVTYPQ	THLHV	YAAK	DIFDVQGHHPNIERSK	GPPLTGKT	VIDDI	WLNN
	FaGmV-20	KT862243	TAATATTAT	LLTYAQ	IHLHA	YAIK	DIFDVDGHHPNISQSR	GRSRTGKT	IFDDI	WLAN
	FaGmV-20	KT862244	TAATATTAT	LLTYAQ	IHLHA	YAIK	DIFDVDGHHPNISQSR	GRSRTGKT	IFDDI	WLAN
	FaGmV-21	KT862245	TAATATTAT	LLTYAQ	THLHV	YAIK	DVFDVLGYHPNIEPSR	GESRTGKT	IFDDI	WISN
	FaGmV-20	KT862246	TAATATTAT	LLTYAQ	IHLHA	YAIK	DIFDVDGHHPNISQSR	GRSRTGKT	IFDDI	WLAN
	FaGmV-21	KT862247	TAATATTAT	LLTYAQ	THLHV	YAIK	DVFDVLGYHPNIEPSR	GESRTGKT	IFDDI	WISN
	FaGmV-18	KT862248	TAATACTAT	LLTYSQ	THLHV	YAVK	GFFDVGGKHPNVVPSR	GPSRLGKT	VFDDL	WLAN
	FaGmV-16	KT862249	TAATATTAT	LLTYAQ	VHLHV	YAIK	DIFDVEGRHPNVVPSK	GPSRTGKT	VFDDM	WCAN
	FaGmV-19	KT862250	TAATGTTAT	LLTYPQ	THLHV	YATK	DYFDVGGKHPNVVPSK	GPSRLGKT	VFDDM	WLAN
	FaGmV-16	KT862251	TAATATTAT	LLTYAQ	VHLHV	YAIK	DIFDVEGRHPNVVPSK	GPSRTGKT	VFDDM	WCAN
	SaGmV-3	KT862252	TAATGTTAT	LLTYAQ	LHIHA	YAIK	RVFDMDGCHPNIVRGY	GPTKLGKT	VFDDM	YLYN
	FaGmV-22	KT862253	TAATGCTAT	LITYAQ	IHLHC	YAIK	DIFDVDGRHPNIAPSY	GDSRVGKT	VFDDI	WISN
	FaGmV-15	KT862254	TATAAATAC	LLTYSQ	IHLHA	YAVK	RLFDVNGQHPNITPCG	GESKLGKT	IFDDI	WLTN
	SaGmV-3	KT862255	TAATGTTAT	LLTYAQ	LHIHA	YAIK	RVFDMDGCHPNIVRGY	GPTKLGKT	VFDDM	YLYN
Smacoviruses	BofSmV-2	KT862218	TAGTGTTAC	MMTGPR	EHWQI	YERK	-	PVGARGKS	IIIDI	VFTN
	ShfSmV-3	KT862219	AAGTGTTAC	MLTMPW	RHIHV	YEKK	-	ETGSIGKS	LIIDI	VLTN
	ShfSmV-1	KT862220	TAGTGTTAC	VLTVPR	RHYQI	YERK	-	DGGWAGKS	IVIDL	VFSN
	ShfSmV-2	KT862221	TAGTGTTAC	MMTIPR	DHFQI	YERK	-	PVGARGKS	IIIDI	VFTN
	BofSmV-3	KT862222	ATTTCTTAC	SVTIPR	HHYQC	YCRK	-	TKGGSGKT	IWIDL	VTTN
	BofSmV-1	KT862223	CACTGTTAC	DAFFPE	EHFQC	YVQK	-	DKGRHGKS	YIIDM	VFCN
	BofSmV-5	KT862224	CACTGTTAC	ERALPD	THYQC	YEEK	-	KQGGNGKS	YVFDL	VFTN
	PofSmV-1	KT862225	TAGTGTTAC	VLTIPR	EHFQI	YERK	-	DKGSIGKS	IVVDL	VTSN
	PoSCV-1	KT862226	TAGTGTTAC	MMTIPR	EHWQI	YETK	-	ETGNVGKS	VIIDV	VMTN
	PoSCV-1	KT862227	TAGTGTTAC	MMTIPR	EHWQI	YETK	-	ETGNVGKS	VIIDI	VMTN
	BofSmV-4	KT862228	TAGTGTTAC	VMTIPQ	RHYQV	YELK	-	HVGGKGKT	VIIDC	ILSN
	BofSmV-6	KT862229	GAGTGTTAC	DVTAPQ	KHWQI	YIMK	-	PLGGIGKT	FIIDV	VLTN
Unclassified CRESS DNA viruses	BofCV-1	KT862230	TAGTATTAC	CFTINN	PHWQG	YCKK	-	GPPGTGKS	VFEEF	IITS
	DufCV-1	KT862231	TAATACTAG	QINSKS	LHLHA	YVMK	-	APPGMGKT	VLDEF	IIMS
	DefCV-1	KT862232	TACTATACC	FLTFPQ	PHWQG	YCTK	-	GETGSGKS	LFDDF	FFTS
	DufCV-2	KT862233	TAATATTAA	LLTYPQ	EHYHA	YVKK	-	GKTGTGKT	IFDDV	IFTS
	DufCV-3	KT862234	TAATATTAA	LITYPQ	EHYHA	YVKK	-	GPTGTGKT	IFDDI	IFTS
	LlfCV-1	KT862235	TAGTATTAC	CFTLNN	PHLQG	YCSK	-	GSPGVGKS	IIDDF	IVTS
	ChfCV-1	KT862236	GGTATATTC	-	PVYHP	YISK	-	APPNSGKT	VYDDR	IVLS
	BofCV-2	KT862237	TATTAATAC	LLTYPR	EHIHV	YVKK	-	GSTGTGKT	ILDDI	YITA
Circular molecules	PofCM-1	KT862256	TACTATTAC	TMTVKN	QHCHF	YLNK	-	GPSGSGKS	WFDEF	ISTV
	BofCM-1	KT862257	TGTAATTAA	CFTLNN	PHLQG	YCSK	-	GLPGVGKS	IIDDF	IVTS

Figure 2.3: Maximum-likelihood phylogenetic trees of the Rep and CP amino acid sequences of gemycircularviruses. The tree is rooted with geminivirus sequences. For each gemycircularvirus, sample type is indicated by a colour coding system and isolation source is indicated by a two-letter code. The gemycircularviruses recovered in this study are shown in bold.



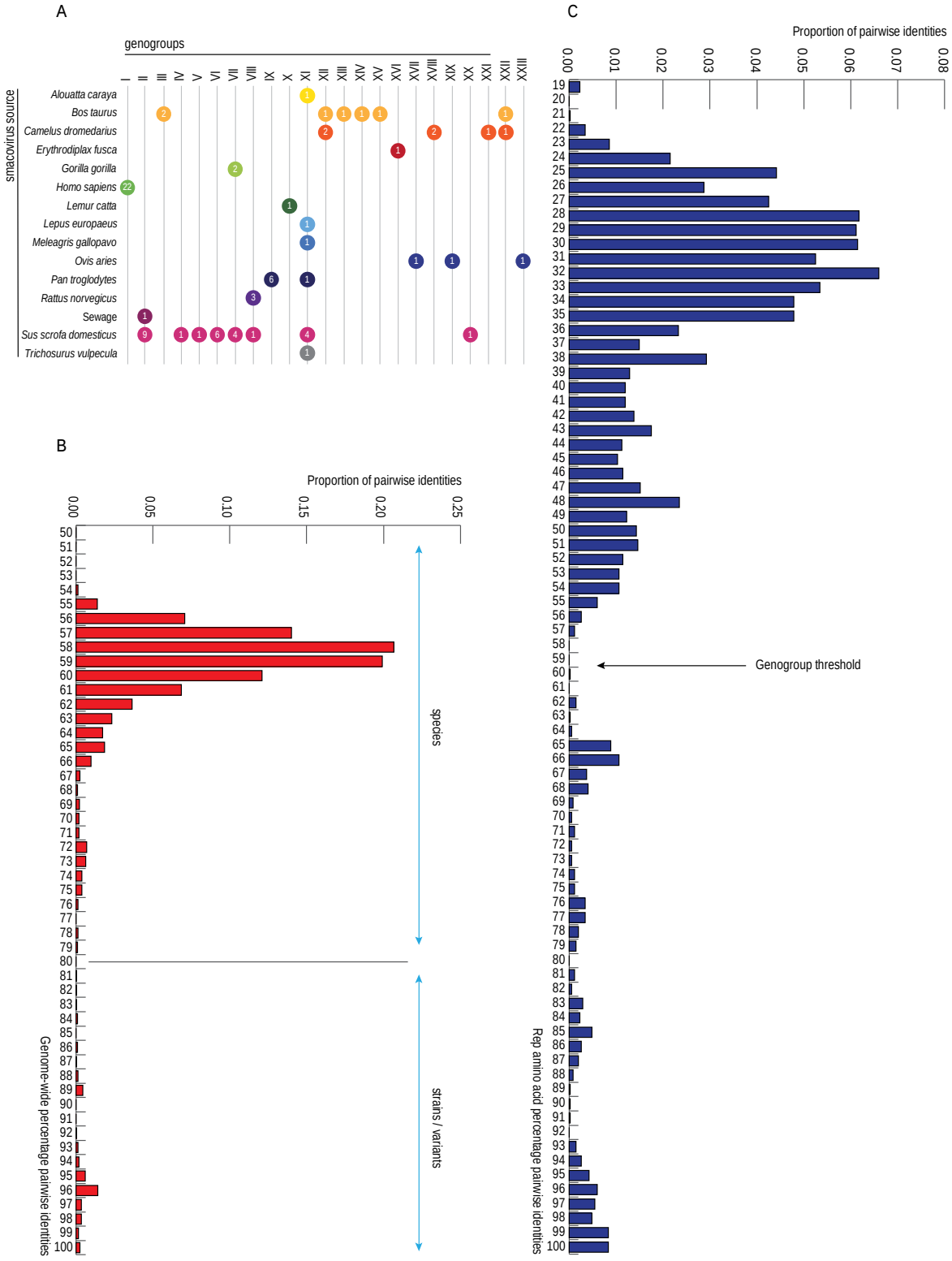
2.4.3 Smacoviruses

Smacoviruses have predominantly been found associated with animal faecal matter, i.e. *Alouatta caraya*, (n=1), *Gorilla gorilla* (n=2), *Lemur catta* (n=1), *Pan troglodytes* (n=7), *B. taurus* (n=1), *Camelus dromedarius* (n=6), *Meleagris gallopavo* (n=1), *Rattus norvegicus* (n=3) and *S. scrofa domesticus* (n=26), and human faeces (n=22). Additionally, smacoviruses have been identified in a sewage oxidation pond sample (n=1), *Erythrodiplax fusca* abdomen (n=1) (Table 2.3). Smacoviruses have two main ORFs, encoding the Rep and CP. In most smacoviruses these ORFs are bidirectionally organised, however, a small group recovered from *S. scrofa domesticus* faces (Pig stool associated circular ssDNA virus, PigSCV; JQ023166, JX305991 - JX305998) and a sewage oxidation pond sample (Sewage-associated circular virus-9, SaCV-9; KJ547633) have a unidirectional genome organisation (Table 2.3).

In this study, twelve smacoviruses were identified in the faecal samples of three *B. taurus* (n=6) and one each of *L. europaeus* (n=1), *O. aries* (n=3), *S. scrofa domesticus* (n=1) and *T. vulpecula* (n=1) (Figure 2.1; Table 2.4). For the purpose of this study and towards setting some basic guidelines for the classification of smacoviruses, the distribution of pairwise identities of the genomes of ambisense smacoviruses was determined using SDT v1.2 (Muhire *et al.*, 2014). Based on the distribution of pairwise identities (Figure 2.4), an 80% genome-wide pairwise identity was used as a species delineator, i.e. smacovirus sequences with <80% genome-wide pairwise identity to previously identified smacoviruses were assigned as a tentative new species.

Two isolates from *L. europaeus* and *T. vulpecula* share >93% genome-wide pairwise identity with Porcine stool-associated circular virus (PoSCV) isolates (JX274036, KF193403) from New Zealand and South Korea, thus these have been named PoSCV (KT862226, KT862227). The four PoSCVs share ~73% genome-wide pairwise identities with PoSCV-1 (KJ577810 - KJ577811) from the USA. The remaining ten smacoviruses from this study share <80% pairwise identity to other smacoviruses and have been assigned the following names based on the source of the faecal sample: Bovine faeces associated smacovirus (BofSmV) -1, -2, -3, -4, -5 and -6, Porcine faeces associated smacovirus-1 (PofSmV-1) and Sheep faeces associated smacovirus (ShfSmV) -1, -2 and -3 (Figure 2.1; Table 2.4). The ambisense smacoviruses have ~46% diversity (Figure 2.4). The new smacoviruses identified in this study share

Figure 2.4: a) Distribution of smacoviruses recovered from faecal sources across smacovirus genogroups I- XXIII. **b)** Distribution of genome-wide pairwise identity scores. The 80% pairwise identity species cut-off applied to smacovirus isolates is shown. **c)** Distribution of Rep amino acid pairwise identity scores. The 60% threshold applied for genogrouping is shown.



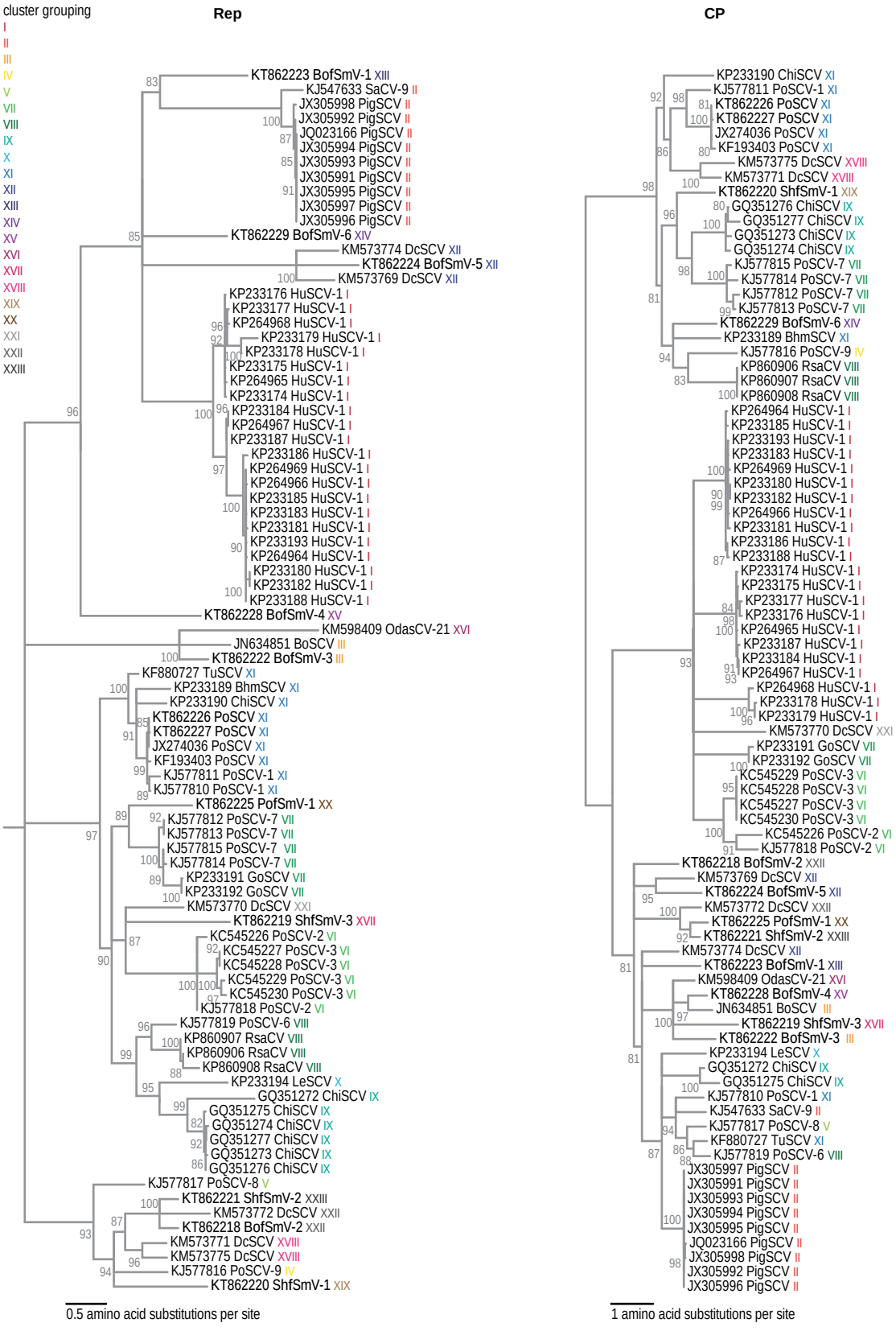
between 55-73% genome-wide identity with previously identified smacoviruses, suggesting that there is significant diversity amongst these viruses. In all smacoviruses from this study, 'NANTGTTAC' was identified as the nonanucleotide motif, with the exception of BofSmV-3 (KT862222) where the putative motif is 'ATTTCTTAC' (Table 2.5).

The recovered smacoviruses have ~2.4 - 2.6kb circular ssDNA genomes that encode the Rep ORF (735 - 822nts) on the complementary-sense strand and the CP ORF (954 - 1113nts) on the virion-sense strand (Figure 2.2). All smacoviruses recovered in this study have a type IV genome organisation (Table 1.2) (Rosario *et al.*, 2012b). The ORFs do not overlap, therefore the smacovirus genomes have two intergenic regions, named the long intergenic region (LIR) and short intergenic region (SIR). Both ORFs are transcribed towards the origin of replication, which is located in the LIR of ShfSmV-2, PofSmV-1, PoSCV and BoFSmV-6, and the SIR of the remaining isolates. The nonanucleotide motif is located on the virion-sense strand in a characteristic stem-loop structure at the origin of replication.

'Genogroups' (A-K) have been proposed by Ng *et al.* (2015) based on Rep amino acid sequence identity. The Reps (n=84) of all smacovirus sequences available in GenBank were analysed, including those from this study (Table 2.3), where it was found that the 60% threshold for genogrouping holds (Figure 2.4). However, based on the criteria proposed (Ng *et al.*, 2015), 23 genogroups were identified. This study opted to assign genogroups to the smacoviruses with Roman numerals (I – XXIII; Table 2.3) rather than letter of the alphabet to allow for expansion of the group, which is likely to occur with significant activity in viral metagenomics. The smacoviruses from this study fall into eleven genogroups (III, XI, XII, XIII, XIV, XIX, XV, XVII, XX, XXII and XXIII), three to previously assigned genogroups III (C) and XI (K) by (Ng *et al.*, 2015) and nine smacoviruses into nine new genogroups (Figures 2.4 and 2.5).

Some of the isolates identified in this study belong to genogroups containing smacoviruses from the faeces of multiple species (Figure 2.4). The two PoSCV isolates belong to genogroup XI, which also contains smacoviruses recovered from the faeces of *A. caraya* (n=1), *L. europaeus* (n=1), *M. gallopavo* (n=1), *P. troglodytes* (n=1), *T. vulpecula* (n=1) and *S. scrofa domesticus* (n=2). BofSmV-5 is classified in genogroup XII with two of the six

Figure 2.5: Maximum-likelihood phylogenetic trees of the Rep and CP amino acid sequences of smacoviruses. The tree is rooted with the Rep and CP sequences of Avon-Heathcote Estuary associated circular virus 29 isolates. Colour-coded Roman numerals indicate the genogroup of each smacovirus. The smacoviruses recovered in this study are shown in bold.



smacoviruses identified in *C. dromedarius* faeces. One of the *C. dromedarius* smacoviruses is classified in genogroup XXII, which also contains BofSmV-2. Other genogroups are species-specific at this stage (Figure 2.4). Genogroup III contains BofSmV-3 and another smacovirus isolated from *B. taurus* faeces, BoSCV (JN634851) while BofSmV-1, BofSmV-6, ShfSmV-3, ShfSmV-1, BofSmV-4, PoSmV-1 and ShfSmV-2 are the only members of genogroups XIII, XIV, XVII, XIX, XV, XX and XXIII, respectively.

There is high diversity in the Reps of smacoviruses (~80%). The genogroups based on percentage pairwise identities of the Rep amino acid sequences are well supported in the Rep maximum-likelihood phylogenetic tree (Figure 2.5). In the maximum-likelihood phylogenetic tree of smacovirus CP sequences, it is evident that certain genogroups are not monophyletic. For example, the CPs of genogroups I, VI, VII (GoSCV isolates) and XXI are more closely related to each other sharing >42% pairwise amino acid identity whereas their Reps share >28% pairwise amino acid identity. Even though the Reps of genogroup VII are closely related to each other sharing >81% pairwise amino acid identity, the CPs of isolates from *G. gorilla* and *S. scrofa domesticus* are significantly different sharing 24-28% pairwise amino acid identity (Figure 2.5).

RCR motifs I, II and III are conserved in the Reps of recovered smacoviruses (Table 2.5). RCR motif I was identified as (M/L/V)Tx(P/S)(R/W/A) in the majority of isolates, where 'x' denotes any amino acid residue. The sequence of RCR motif I differed in BofSmV-1 and BofSmV-5, whose sequences are MGFIT and ERALP, respectively. The sequence of RCR motif II is Hx(Q/R)(I/V/C)(R/A/K)x. Like recovered gemycircularvirus Reps, the smacovirus Reps contain a single tyrosine residue in RCR motif III (YxxK) so are classified as superfamily II Reps. The predominant sequence is YERK which is found in BofSmV-2, ShfSmV-1, ShfSmV-2 and PofSmV-1. The sequence of the Walker A motif in recovered smacovirus isolates is xxGxxGK(S/T) (Table 2.5) and the Walker B sequence is (I/W/V)(I/L/V)DxP, except for BofSmV-6 whose Walker B motif is FIIDV. Motif C was identified as (V/I)x(T/S/C)N. Importantly, an invariant aspartic acid residue was identified in the Walker B motifs, demonstrating that the putative Reps are likely capable of hydrolysing ATP to mediate helicase activity.

2.4.4 Unclassified CRESS DNA viruses

Eight CRESS DNA viruses were identified that do not fall within any of the known groupings (Figure 2.1; Table 2.4). The Reps of these viruses share <75% identity, hence they have simply been named based on the faecal source: Bovine faeces associated circular DNA virus-1 and -2 (BofCV-1, -2) from *B. taurus*, Duck faeces associated circular DNA virus-1, -2 and -3 (DufCV-1, -2, -3) from *A. platyrhynchos*, Deer faeces associated circular DNA virus-1 (DefCV-1) from *D. dama*, Llama faeces associated circular DNA virus-1 (LlfCV-1) from *L. glama* and Chamois faeces associated circular DNA virus-1 from *R. rupicapra*. DufCV-2 and -3 (KT862233 - KT862234) were recovered from the same *A. platyrhynchos* sample.

The unclassified CRESS DNA viruses identified in this study have circular ssDNA genomes ranging in size from ~1.8 – 3.0kb with varying genome architectures (Figure 2.2; Table 1.2). The nonanucleotide motif (Table 2.5) of most isolates is located in the short intergenic region (SIR) on the Rep-encoding strand, except for BofCV-2 whose origin of replication is located in the long intergenic region (LIR) and ChfCV-1 which contains a single IR. The Rep ORF ranges in size from 750 - 1236nts and the CP ORF ranges from 513 - 1116nts. DefCV-1, DufCV-2 and DufCV-3 have a type I circular ssDNA genome architecture i.e. the Rep and CP ORFs are located on the virion-sense and complementary-sense strands, respectively, and are both transcribed away from the putative origin of replication. The genome of BofCV-2 differs in that the Rep and CP ORFs are transcribed towards the origin of replication, therefore BofCV-2 has a type III circular ssDNA genome architecture. BofCV-1 has a type IV genome organisation where the ORFs are transcribed towards the origin of replication and the CP and Rep ORFs are located on the virion-sense and complementary-sense strands, respectively. DufCV-1, LlfCV-1 and ChfCV-1 have a unisense ORF orientation where the Rep and CP ORFs are located on the virion-sense strand. Hence, these viruses have a type V circular ssDNA genome architecture.

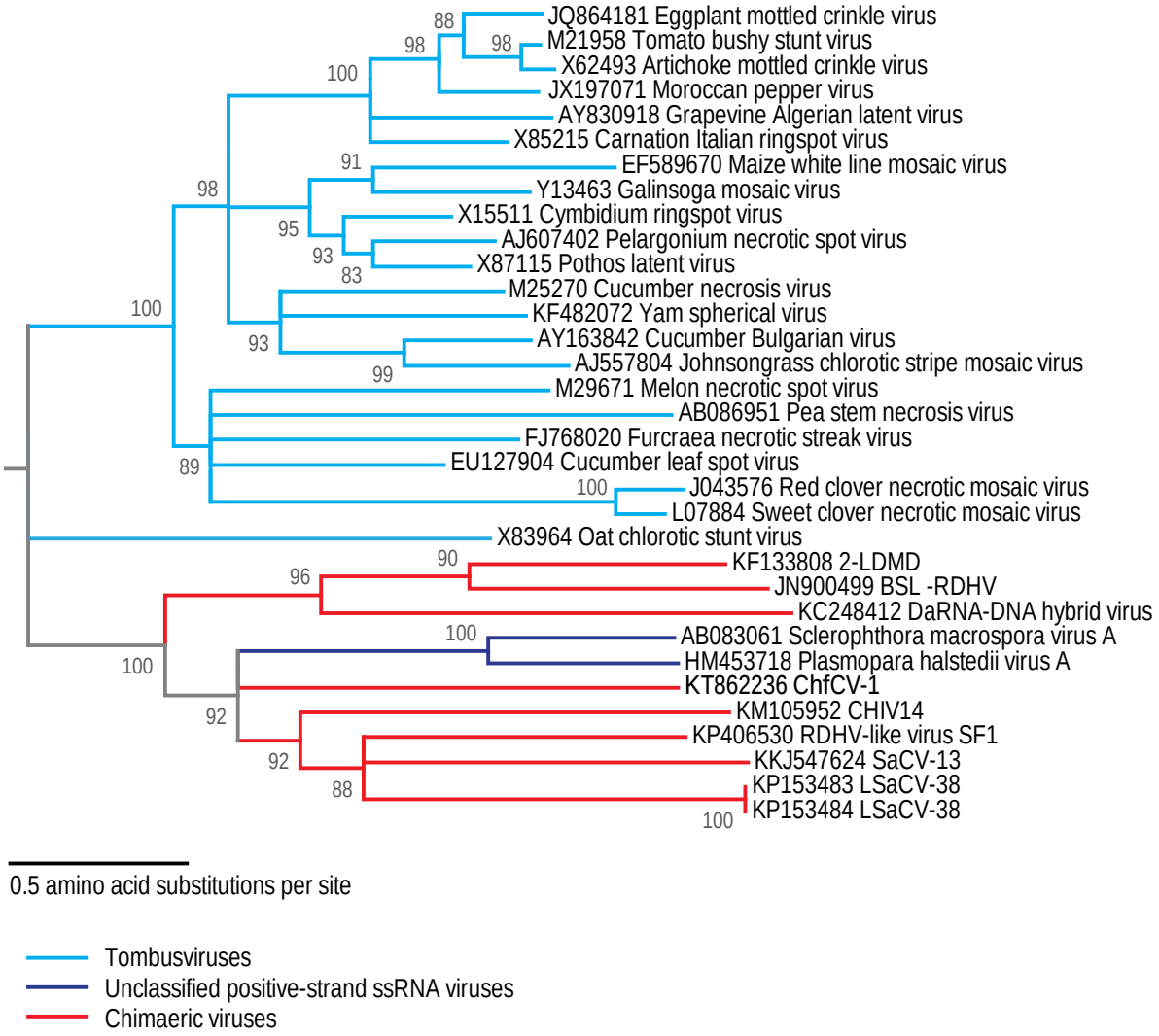
In each of these circular molecules, a putative CP and Rep were identified. The BLASTp analysis of the putative Rep and CPs encoded by these CRESS DNA viruses is summarised in Table 2.6. BofCV-1 is most closely related to Dromedary stool-associated circular ssDNA virus (KM573765) identified in faeces of *C. dromedarius* from the United Arab Emirates (Rep: 60% identity, coverage 92%; e-value $5x10^{-148}$). DufCV-1 is most closely related

Table 2.6: Top hit identified from BLASTp analysis of the Rep and CP amino acid sequences of unclassified CRESS DNA viruses and circular molecules identified in this study.

		ORF	BLASTp hit	Accession #	% Pairwise identity	E-value	Query coverage	Isolation source	Country
KT862230	BofCV-1	Rep	Dromedary stool-associated circular ssDNA virus	KM573765	60	5e-148	92	*Camelus dromedarius*	UAE
		CP	Dromedary stool-associated circular ssDNA virus	KM573765	45	2e-28	51	*Camelus dromedarius*	UAE
KT862231	DufCV-1	Rep	Sewage-associated circular DNA virus-30	KM821765	48	4e-138	99	Sewage	New Zealand
		CP	Sewage-associated circular DNA virus-30	KM821765	38	4e-25	68	Sewage	New Zealand
KT862232	DefCV-1	Rep	Acheta domesticus volvovirus	KC794540	41	1e-58	77	Acheta domesticus	Japan
		CP	Acheta domesticus volvovirus	KC794540	41	1e-18	25	Acheta domesticus	Japan
KT862233	DufCV-2	Rep	Dromedary stool-associated circular ssDNA virus	KM573764	36	2e-45	96	*Camelus dromedarius*	UAE
		CP	Odonata-associated circular virus-14	KM598397	32	6e-11	86	*Erythrodiplax fusca*	USA
KT862234	DufCV-3	Rep	Dromedary stool-associated circular ssDNA virus	KM573764	34	3e-39	92	*Camelus dromedarius*	UAE
		CP	Odonata-associated circular virus-14	KM598397	29	7e-10	80	*Erythrodiplax fusca*	USA
KT862235	LlfCV-1	Rep	Bat circovirus*	KJ641730	69	1e-142	100	*Murina leucogaster*	China
		CP	Bat circovirus*	KJ641730	40	1e--56	97	*Murina leucogaster*	China
KT862236	ChfCV-1	Rep	Odonata-associated circular virus-8	KM598391	30	9e-19	73	Libellula quadrimaculata	USA
		CP	RDHV-like virus SF1	KP406530	45	1e-38	52	Water	USA
KT862237	BofCV-2	Rep	Dromedary stool-associated circular ssDNA virus	KM573764	37	3e-46	93	*Camelus dromedarius*	UAE
		CP	-	-					
KT862256	PofCM-1	Rep	Fur seal faeces associated circular DNA virus	KF246569	64	2e-171	89	*Arctocephalus forsteri*	New Zealand
KT862257	BofCM-1	Rep	Sewage-associated circular DNA virus-17	KM821752	66	2e-140	100	Sewage	New Zealand

* Not circovirus but unclassified CRESS DNA virus

Figure 2.6: Maximum-likelihood phylogenetic tree of the CP amino acid sequences of ChfCV-1, RNA-DNA hybrid viruses and representative tombusviruses. A colour-coding system indicates if the virus is a tombusvirus (light blue), unclassified positive-strand ssRNA virus (dark blue) or chimaeric virus (red). The ChfCV-1 isolate recovered in this study is shown in bold.



to Sewage-associated circular DNA virus-30 (KM821765) whose Rep shares 48% identity (coverage 99%; e-value $4x10^{-138}$), whereas LlfCV-1 is most closely related to an unclassified CRESS DNA virus from bats (Bat circovirus, KJ641730) whose Rep shares 59% identity (coverage 100%; e-value $1x10^{-142}$).

ChfCV-1 encodes a CP that shares 45% identity (coverage 52%; e-value $1x10^{-38}$) with that of RDHV-like virus SF1 (KP406530) from a water sample in the USA, however, the CP of this virus is most closely related to CPs encoded by RNA viruses in the *Tombusviridae* family (Figure 2.6; Table 2.6) (Dayaram *et al.*, 2016; Diemer & Stedman, 2012; Krupovic *et al.*, 2015; Roux *et al.*, 2013).

SsDNA viruses frequently undergo recombination with genomes replicating within the same host cell (Martin *et al.*, 2011a). This allows rapid exploration of viral sequence space and can give rise to natural chimaeras when horizontal gene transfer events occur between co-replicating, divergent viruses. A RNA-DNA hybrid virus (RDHV) was the first natural chimaera of RNA and DNA viruses to be described (Diemer & Stedman, 2012). This virus was isolated from Boiling Springs Lake, USA and encodes a circovirus-like Rep and a CP most similar to unclassified single-stranded RNA (ssRNA) viruses and tombusviruses. Other hybrid viruses have been identified in contaminated spin-columns (Krupovic *et al.*, 2015), a sewage oxidation pond sample (Kraberger *et al.*, 2015a), fresh and marine water bodies (Dayaram *et al.*, 2016; Hewson *et al.*, 2013b; McDaniel *et al.*, 2014), wastewater (Greninger & DeRisi, 2015) and a dragonfly (Rosario *et al.*, 2012a).

2.4.5 Circular molecules

Two circular DNA molecules (1285 - 1479nts) were recovered from *B. taurus* and *S. scrofa domesticus* and have been named Bovine faeces associated circular DNA molecule-1 (BofCM-1) and Porcine faeces associated circular DNA molecule-1 (PofCM-1). Both BofCM-1 and PofCM-1 encode a Rep that is most closely related to that of Fur seal faeces associated circular DNA virus (KF246569) and Sewage-associated circular DNA virus-17 (KM821752), respectively (Table 2.6). The Rep ORF of BofCM-1 is 843nts and the Rep ORF of PofCM-1 is 1206nts in size. The

nonanucleotide motif of both circular molecules is located on the Rep-encoding strand, therefore the circular molecules have a type VII genome architecture (Figure 2.2; Table 1.2) (Rosario *et al.*, 2012b). These molecules may be genome segments of multicomponent CRESS DNA viruses like nanoviruses or geminiviruses, or satellite molecules similar to those associated with begomoviruses (*Geminiviridae* family) (Briddon *et al.*, 2003; Horser *et al.*, 2001b; King *et al.*, 2011; Kumar *et al.*, 2013; Patil & Fauquet, 2010; Rohde *et al.*, 1990; Saunders *et al.*, 2000). It is also likely that these molecules could be sub-genomic defective molecules that are produced during rolling circle replication of the virus (Jeske *et al.*, 2001; Martin *et al.*, 2011a; van der Walt *et al.*, 2009).

2.5 Conclusion

Up until 2013, the only CRESS DNA viruses identified in New Zealand were two circoviruses, *Beak and feather disease virus* (BFDV) (Ha *et al.*, 2009; Ha *et al.*, 2007; Jackson *et al.*, 2014a; Jackson *et al.*, 2015; Massaro *et al.*, 2012; Ortiz-Catedral *et al.*, 2010; Ritchie *et al.*, 2003) and *Porcine circovirus-2* (PCV-2) (Garkavenko *et al.*, 2005; Neumann *et al.*, 2007), and two begomoviruses, *Abutilon mosaic viru*s and *Honeysuckle yellow vein virus* (Lyttle & Guy, 2004). However, in the last three years, a large number of novel CRESS DNA viruses have been identified in a freshwater lake (Dayaram *et al.*, 2016), insects (Dayaram *et al.*, 2014; Dayaram *et al.*, 2016; Dayaram *et al.*, 2013c), molluscs (Dayaram *et al.*, 2015a; Dayaram *et al.*, 2016; Dayaram *et al.*, 2013a; b), sediment (Dayaram *et al.*, 2015a; Dayaram *et al.*, 2016; Kraberger *et al.*, 2013), a sewage oxidation pond (Kraberger *et al.*, 2015a), plant material (Kraberger *et al.*, 2015b), parrot nesting material (Sikorski *et al.*, 2013c) and animal faecal matter (Sikorski *et al.*, 2013a; Sikorski *et al.*, 2013b; Sikorski *et al.*, 2013d) through various approaches involving RCA and viral metagenomics (Table 1.3).

Fourteen gemycircularviruses (Sikorski *et al.*, 2013d), a smacovirus (Sikorski *et al.*, 2013a) and an unclassified CRESS DNA virus (Sikorski *et al.*, 2013b) have previously been identified from animal faecal matter in New Zealand. This

dissertation expanded upon this baseline data using viral metagenomic approaches. Through a next-generation sequencing-informed approach, 38 CRESS DNA viruses and two CRESS DNA molecules were identified. Eighteen of the identified CRESS DNA viruses were classified as gemycircularviruses, which represented eleven species, nine of which are novel. Twelve of the CRESS DNA virus isolates displayed significant relatedness to smacoviruses. Based on genome-wide nucleotide pairwise identities, this constituted eleven smacovirus species, including ten novel species. The remaining CRESS DNA viruses are unclassified.

Overall, the faecal matter of wild and domestic animals has proved to be an effective tool for characterising novel CRESS DNA viruses in New Zealand. In addition, some of the recovered CRESS DNA viruses share a high degree of sequence similarity with gemycircularviruses identified previously in New Zealand from the faeces of a different animal, such that they were assigned as members of the same species. This reinforces the idea that CRESS DNA viruses are prevalent in nature and have a broad distribution.

This study contributes significantly to our knowledge of CRESS DNA virus diversity, specifically of CRESS DNA viruses circulating within animal faecal matter in New Zealand. The CRESS DNA viruses described in this study were recovered from nineteen faecal samples of twelve animal species collected across the South Island of New Zealand. The CRESS DNA viruses identified show a wide diversity, consisting of members belonging to two large CRESS DNA viral groups and a collection of divergent CRESS DNA viruses.

Chapter Three: Concluding Remarks

The number of ssDNA viruses recorded in the literature has expanded rapidly in recent years. Viral metagenomic approaches have largely been used to explore the global ssDNA viral sequence space. Innovations in next-generation sequencing technologies have generated more accurate platforms that are cheaper, quicker and support longer reads. Such approaches have overcome biases in previous strategies where the majority of research focussed on viruses with economic or medical relevance that were easily cultured. The use of Phi29 DNA polymerase coupled with next-generation sequencing platforms has successfully identified CRESS DNA viruses from a range of sources, including many environmental samples. Formally classified CRESS DNA viruses, i.e. circoviruses, geminiviruses and nanoviruses, are well characterised while other CRESS DNA viruses are found to have divergent genomes to the extent that they cannot be classified into existing taxa. The diverse collection of CRESS DNA viruses uncovered through metagenomic studies highlights inadequacies in the current viral taxonomic system and a large gap in knowledge of CRESS DNA viruses and the wider global virome. Accordingly, several research groups have dedicated their time to unravelling the true diversity and prevalence of CRESS DNA viruses.

3.1 Main findings

This dissertation presents the complete genomes of 38 CRESS DNA viruses and two circular molecules from 49 animal faecal samples collected from sites across the South Island of New Zealand. A large proportion of isolates shared similarities with proposed CRESS DNA virus groupings, namely the gemycircularviruses and smacoviruses. Eighteen CRESS DNA virus isolates recovered from twelve faecal samples were assigned as gemycircularviruses, which adds considerably to the number of gemycircularviruses identified in New Zealand and provides more support for this proposed grouping. The recovered gemycircularvirus isolates are diverse, with fourteen isolates assigned to nine novel species and two previously identified gemycircularvirus species. Three isolates were assigned to SaGmV-3 (KJ547643) and

one isolate was assigned to FaGmV-4 (KF371638), which were also identified from faecal sources collected in the South Island of New Zealand. Taking the gemycircularviruses recovered in this study into account, 43% of gemycircularviruses were identified in New Zealand. As the first gemycircularvirus was not identified in New Zealand until 2013, this demonstrates the success of viral metagenomic studies conducted by a single research group in filling this gap in knowledge over the last few years. Previous studies have seen the identification of CRESS DNA virus species, namely SsHADV-1 and *Starling circovirus* isolates, in multiple countries (Dayaram *et al.*, 2013a; Dayaram *et al.*, 2015b; Kraberger *et al.*, 2013). Therefore the discovery of FaGmV-14 isolates that share significant sequence similarity with MmFV (JN704610), a virus first discovered in *M. meles* faeces in the Netherlands, further supports the notion that gemycircularviruses and other CRESS DNA viruses may have a wider distribution than first thought.

Twelve CRESS DNA virus isolates were identified from seven faecal samples that share significant sequence similarities with smacoviruses. The isolates were divided into ten novel smacovirus species and one previously described smacovirus species, PoSCV, based on genome-wide nucleotide pairwise identities. PoSCV and SaCV-9 were the only smacoviruses identified in New Zealand prior to this study (Kraberger *et al.*, 2015a; Sikorski *et al.*, 2013a). This thesis research therefore presents the first baseline data on the diversity of smacoviruses circulating within New Zealand. In addition, a framework for the classification of smacoviruses into 23 genogroups was presented, as adapted from Ng *et al.* (2015). New Zealand smacoviruses are classified into twelve genogroups. Five of the genogroups contain smacoviruses recovered in other countries (Germany, China, South Korea, Hungary, the USA and UAE), while the remaining seven genogroups only contain smacoviruses identified in New Zealand. This demonstrates that there is significant diversity within New Zealand smacoviruses and within members of the proposed smacovirus grouping as a whole.

Two circular molecules and the remaining eight CRESS DNA viruses identified in this study are highly divergent and remain unclassified. This emphasises that the current viral taxonomic scheme does not encompass the true diversity of the global virome. The circular molecules may represent single components of a multicomponent virus or defective molecules, as identified in other studies conducted

in New Zealand and the Kingdom of Tonga (Kraberger *et al.*, 2015a; Male *et al.*, 2016; Stainton *et al.*, 2016). BLASTp analysis of the unclassified CRESS DNA virus Rep and CP ORFs showed the highest similarities to ssDNA viruses, with the exception of ChfCV-1 whose CP was most similar to a chimaeric virus, RDHV-like virus SF1 (KP406530). The capture of CP sequences from co-infecting ssRNA viruses is encouraged by the propensity of ssDNA virus genomes to undergo non-homologous recombination. Stedman (2013) speculates that the incorporation of messenger RNA (mRNA) into the ssDNA viral genome is most likely to occur during priming or ligation events. This process does not preclude integration of Rep mRNA into ssDNA virus genomes, however the higher abundance of CP mRNA compared to Rep mRNA in RNA viruses and the role of the CP in virus transmission favours its maintenance in RNA-DNA hybrid viruses after selection (Stedman, 2013). Non-homologous recombination is a known gene acquisition mechanism in dsDNA viruses, providing support for a similar process that mediates RNA capture by ssDNA viruses (Hendrix *et al.*, 2000; Stedman, 2013). A number of ssDNA viruses encoding a CP related to unclassified ssRNA viruses have been identified in the USA (Diemer & Stedman, 2012; Greninger & DeRisi, 2015; Hewson *et al.*, 2013b; Krupovic *et al.*, 2015; McDaniel *et al.*, 2014; Rosario *et al.*, 2012a) and New Zealand (Dayaram *et al.*, 2016; Kraberger *et al.*, 2015a). This suggests that RNA-DNA mosaicism may be commonplace in nature. Natural chimeras of RNA and DNA viruses are of particular importance as they may provide insight into the transition from RNA-based genomes to DNA-based genomes.

The research presented in this thesis further demonstrates that animal faecal matter is a successful sampling method. A large number of CRESS DNA viruses have been identified in faecal sources (Table 3.1), including 88% of gemycircularviruses identified in New Zealand and 70% of gemycircularviruses discovered globally. In addition, 99% of smacoviruses recovered outside New Zealand and all of the New Zealand smacoviruses were recovered from faecal sources. The use of animal faeces is an appropriate tool as it is non-invasive, supports the growth of varied host species and maintains the stability of viral communities (Abrahão *et al.*, 2009; de Carvalho Ferreira *et al.*, 2014; Ramos *et al.*, 2000). The use of faecal samples in diagnostics is not a new concept and has been applied to the detection of viral pathogens in humans,

including noroviruses (Ambert-Balay & Pothier, 2013; Bavelaar *et al.*, 2015; Khamrin *et al.*, 2011; Kirby & Iturriza-Gómara, 2012; Nakamura *et al.*, 2009).

3.2 Future directions

Metagenomic studies have identified a large number of unclassified CRESS DNA viruses, including 272 unclassified CRESS DNA virus isolates in New Zealand, highlighting our limited understanding of this seemingly prevalent group of viruses. Indeed, further investigation may reveal that these unclassified CRESS DNA viruses represent members of larger groupings that are yet to be explored in viral metagenomic studies to date. The approach applied in this dissertation could be extended to analyse faecal sources from different animal species and from more sites, including the North Island. It would be of particular interest to sample faeces of native and endemic birds on New Zealand's island sanctuaries. These island sanctuaries also contain native and endemic plants and invertebrates that would represent unique host species for CRESS DNA viruses. The habitat is minimally influenced by humans and introduced species, therefore, these areas may represent distinctive viromes compared to the rest of New Zealand. The diversity of New Zealand's island sanctuaries is unexplored, with the exception of *Beak and feather disease virus* isolates (Jackson *et al.*, 2015; Massaro *et al.*, 2012; Ortiz-Catedral *et al.*, 2010) and eight gemycircularviruses (Sikorski *et al.*, 2013d).

Our understanding of CRESS DNA virus diversity in the Pacific Islands excluding New Zealand, is even more limited. Gemycircularviruses, cycloviruses and unclassified CRESS DNA viruses have been recovered from Pacific flying fox faeces (Male *et al.*, 2016), signalgrass and sugarcane leaves (Male *et al.*, 2015) and dragonfly abdomen (Rosario *et al.*, 2012a) collected in Tonga. Additionally, research has identified pathogenic viruses of economic importance, i.e. *Beak and feather disease virus* (Jackson *et al.*, 2014b; Julian *et al.*, 2012) and *Banana bunchy top virus* (Karan *et al.*, 1994; Stainton *et al.*, 2012; Stainton *et al.*, 2015) in New Caledonia, Fiji, Samoa and Tonga. This demonstrates that the global diversity of CRESS DNA viruses, particularly in the South Pacific is not well understood.

Little information is known about the ecological significance of the CRESS DNA viruses that have been identified. While the use of viral metagenomics enables efficient sampling of viruses circulating within an environment, it does not provide information as to the specific source of the viruses. As discussed elsewhere, gemycircularviruses likely infect fungi (Kraberger *et al.*, 2015a; Liu *et al.*, 2011; Rosario *et al.*, 2012a; Sikorski *et al.*, 2013d; Yu *et al.*, 2010). Faecal matter is quickly colonised by fungi, therefore the recovery of eighteen gemycircularviruses in this study supports this hypothesis. Of the CRESS DNA viruses identified, infectivity studies have only been conducted on one gemycircularvirus, SsHADV-1 (Yu *et al.*, 2010). Thus, the hosts of gemycircularviruses, smacoviruses and the unclassified CRESS DNA viruses are unknown at this stage. Future experiments should therefore aim to identify specific hosts. This can be achieved using probes generated in a manner similar that of Ng *et al.* (2014). Potential host tissues can be exposed to high levels of the multimeric viral clones, followed by extraction and PCR amplification of viral DNA using abutting primers that are designed based on complete genomes of CRESS DNA viruses. Application of this procedure to different host species can therefore identify the host range of recovered CRESS DNA viruses.

Additionally, specific probes can be used to investigate the presence of CRESS DNA viruses in tissue or blood samples of diseased animals. If the virus is present in animal tissue, this presents a risk of transmission and pathogenicity in humans. Specific probes therefore represent a cost-effective surveillance method to identify and monitor the presence of pathogenic CRESS DNA viruses in animals and humans. Future studies should also investigate possible applications of these viruses, such as their efficacy as potential biocontrol agents (Yu *et al.*, 2013; Yu *et al.*, 2010). Experimental evidence suggests that SsHADV-1 induces hypovirulence in its fungal host and may be a suitable candidate as a biocontrol agent against *S. sclerotiorum*-induced crop diseases (Yu *et al.*, 2013). A study where archived gemycircularviruses and fungal samples were screened for hypovirulence-associated factors would be beneficial in New Zealand, as diverse fungal pathogens are known to infect economically important trees and grapevines (Crane *et al.*, 2009; Hood *et al.*, 2015; Pathrose *et al.*, 2014; Spiers, 1998; Watt *et al.*, 2011).

Table 3.1: Description of CRESS DNA viruses and circular molecules identified from faecal sources as of the 5th of January 2016

Accession #	CRESS DNA virus / molecule name	Country	Sample type	Isolation source	Common name	Reference
AB937980	Cyclovirus ZM32	Zambia	Faeces	*Mastomys natalensis*	Natal multimammate mouse	(Sasaki *et al.*, 2015)
AB937981	Cyclovirus ZM01	Zambia	Faeces	*Crocidura hirta*	Lesser red musk shrew	(Sasaki *et al.*, 2015)
AB937982	Cyclovirus ZM36a	Zambia	Faeces	*Crocidura hirta*	Lesser red musk shrew	(Sasaki *et al.*, 2015)
AB937983	Cyclovirus ZM38	Zambia	Faeces	*Crocidura hirta*	Lesser red musk shrew	(Sasaki *et al.*, 2015)
AB937984	Cyclovirus ZM41	Zambia	Faeces	*Crocidura hirta*	Lesser red musk shrew	(Sasaki *et al.*, 2015)
AB937985	Cyclovirus ZM50a	Zambia	Faeces	*Crocidura luna*	Lesser red musk shrew	(Sasaki *et al.*, 2015)
AB937986	Cyclovirus ZM54	Zambia	Faeces	*Crocidura hirta*	Lesser red musk shrew	(Sasaki *et al.*, 2015)
AB937987	Cyclovirus ZM62	Zambia	Faeces	*Crocidura hirta*	Lesser red musk shrew	(Sasaki *et al.*, 2015)
GQ351272	Chimpanzee stool associated circular ssDNA virus	Cameroon	Faeces	*Pan troglodytes*	Chimpanzee	(Blinkova *et al.*, 2010)
GQ351273	Chimpanzee stool associated circular ssDNA virus	Tanzania	Faeces	*Pan troglodytes*	Chimpanzee	(Blinkova *et al.*, 2010)
GQ351274	Chimpanzee stool associated circular ssDNA virus	Tanzania	Faeces	*Pan troglodytes*	Chimpanzee	(Blinkova *et al.*, 2010)
GQ351275	Chimpanzee stool associated circular ssDNA virus	Tanzania	Faeces	*Pan troglodytes*	Chimpanzee	(Blinkova *et al.*, 2010)
GQ351276	Chimpanzee stool associated circular ssDNA virus	Tanzania	Faeces	*Pan troglodytes*	Chimpanzee	(Blinkova *et al.*, 2010)
GQ351277	Chimpanzee stool associated circular ssDNA virus	Tanzania	Faeces	*Pan troglodytes*	Chimpanzee	(Blinkova *et al.*, 2010)
GQ351278	Chimpanzee stool associated circular ssDNA virus	Republic of the Congo	Faeces	*Pan troglodytes*	Chimpanzee	(Blinkova *et al.*, 2010)
GQ404844	Cyclovirus PK5006	Pakistan	Faeces	*Homo sapiens*	Human	(Li *et al.*, 2010a)
GQ404845	Cyclovirus PK5034	Pakistan	Faeces	*Homo sapiens*	Human	(Li *et al.*, 2010a)
GQ404846	Cyclovirus PK5222	Pakistan	Faeces	*Homo sapiens*	Human	(Li *et al.*, 2010a)
GQ404847	Cyclovirus PK5510	Pakistan	Faeces	*Homo sapiens*	Human	(Li *et al.*, 2010a)
GQ404848	Cyclovirus PK6197	Pakistan	Faeces	*Homo sapiens*	Human	(Li *et al.*, 2010a)
GQ404849	Cyclovirus Chimp11	Central Africa	Faeces	*Pan troglodytes*	Chimpanzee	(Li *et al.*, 2010a)
GQ404850	Cyclovirus Chimp12	Central Africa	Faeces	*Pan troglodytes*	Chimpanzee	(Li *et al.*, 2010a)
GQ404854	Cyclovirus NG12	Nigeria	Faeces	*Homo sapiens*	Human	(Li *et al.*, 2010a)
GQ404855	Cyclovirus NG14	Nigeria	Faeces	*Homo sapiens*	Human	(Li *et al.*, 2010a)
GQ404856	Human stool-associated circular virus NG13	Nigeria	Faeces	*Homo sapiens*	Human	(Li *et al.*, 2010a)
GQ404857	Cyclovirus TN25	Tunisia	Faeces	*Homo sapiens*	Human	(Li *et al.*, 2010a)
GQ404858	Cyclovirus TN18	Tunisia	Faeces	*Homo sapiens*	Human	(Li *et al.*, 2010a)
HM228874	Bat cyclovirus GF-4c	USA	Faeces	*Antrozous pallidus*	Bat	(Li *et al.*, 2010b)
HM228875	Circoviridae TM-6c	USA	Faeces	*Tadarida brasiliensis*	Bat	(Li *et al.*, 2010b)
JF713716	Po-Circo-like virus 21	USA	Faeces	*Sus scrofa*	Pig	(Shan *et al.*, 2011)
JF713717	Po-Circo-like virus 22	USA	Faeces	*Sus scrofa*	Pig	(Shan *et al.*, 2011)
JF713718	Po-Circo-like virus 41	USA	Faeces	*Sus scrofa*	Pig	(Shan *et al.*, 2011)
JF713719	Po-Circo-like virus 51	USA	Faeces	*Sus scrofa*	Pig	(Shan *et al.*, 2011)
JF755401	Rodent stool-associated circular genome virus	USA	Faeces	Neotoma cinerea	Bushy-tailed woodrat	(Phan *et al.*, 2011)
JF755402	Rodent stool-associated circular genome virus	USA	Faeces	Microtus pennsylvanicus	Meadow vole	(Phan *et al.*, 2011)
JF755403	Rodent stool-associated circular genome virus	USA	Faeces	Microtus pennsylvanicus	Meadow vole	(Phan *et al.*, 2011)
JF755404	Rodent stool-associated circular genome virus	USA	Faeces	*Microtus pennsylvanicus*	Meadow vole	(Phan *et al.*, 2011)
JF755405	Rodent stool-associated circular genome virus	USA	Faeces	*Microtus pennsylvanicus*	Meadow vole	(Phan *et al.*, 2011)
JF755406	Rodent stool-associated circular genome virus	USA	Faeces	*Microtus pennsylvanicus*	Meadow vole	(Phan *et al.*, 2011)
JF755407	Rodent stool-associated circular genome virus	USA	Faeces	*Microtus pennsylvanicus*	Meadow vole	(Phan *et al.*, 2011)
JF755408	Rodent stool-associated circular genome virus	USA	Faeces	*Mus musculus*	House mouse	(Phan *et al.*, 2011)
JF755409	Rodent stool-associated circular genome virus	USA	Faeces	*Mus musculus*	House mouse	(Phan *et al.*, 2011)
JF755410	Rodent stool-associated circular genome virus	USA	Faeces	*Peromyscus truei*	Pinyon mouse	(Phan *et al.*, 2011)
JF755411	Rodent stool-associated circular genome virus	USA	Faeces	*Microtus pennsylvanicus*	Meadow vole	(Phan *et al.*, 2011)

JF755412	Rodent stool-associated circular genome virus	USA	Faeces	*Microtus pennsylvanicus*	Meadow vole	(Phan *et al.*, 2011)
JF755413	Rodent stool-associated circular genome virus	USA	Faeces	*Microtus pennsylvanicus*	Meadow vole	(Phan *et al.*, 2011)
JF755414	Rodent stool-associated circular genome virus	USA	Faeces	*Microtus pennsylvanicus*	Meadow vole	(Phan *et al.*, 2011)
JF755415	Rodent stool-associated circular genome virus	USA	Faeces	*Mus musculus*	House mouse	(Phan *et al.*, 2011)
JF755416	Rodent stool-associated circular genome virus	USA	Faeces	*Microtus pennsylvanicus*	Meadow vole	(Phan *et al.*, 2011)
JF755417	Rodent stool-associated circular genome virus	USA	Faeces	*Microtus pennsylvanicus*	Meadow vole	(Phan *et al.*, 2011)
JF938078	Bat circovirus ZS/China/2011	China	Faeces	-	Bat	(Ge *et al.*, 2011)
JF938079	Bat circovirus ZS/China/2011	China	Faeces	-	Bat	(Ge *et al.*, 2011)
JF938080	Bat circovirus ZS/China/2011	China	Faeces	-	Bat	(Ge *et al.*, 2011)
JF938081	Bat circovirus ZS/China/2011	China	Faeces	-	Bat	(Ge *et al.*, 2011)
JF938082	Bat circovirus ZS/China/2011	China	Faeces	-	Bat	(Ge *et al.*, 2011)
JN377562	Bat circovirus ZS/Yunnan-China/2009	China	Faeces	-	Bat	(Ge *et al.*, 2011)
JN377566	Bat circovirus ZS/Yunnan-China/2009	China	Faeces	-	Bat	(Ge *et al.*, 2011)
JN377580	Bat circovirus ZS/Yunnan-China/2009	China	Faeces	-	Bat	(Ge *et al.*, 2011)
JN634851	Bovine stool-associated circular virus	South Korea	Faeces	*Bos taurus*	Cow	(Kim *et al.*, 2012)
JN704610	Meles meles fecal virus	Netherlands	Rectal swab	*Meles meles*	European badger	(van den Brand *et al.*, 2012)
JN857329	Circoviridae batCV-SC703	China	Faeces	-	Bat	(Ge *et al.*, 2012)
JQ023166	Pig stool associated circular ssDNA virus GER2011	Germany	Faeces	*Sus scrofa*	Pig	(Sachsenröder *et al.*, 2012)
JQ898331	Baminivirus	Thailand	Sewage	Untreated sewage	-	(Ng *et al.*, 2012)
JQ898332	Niminivirus	Nigeria	Sewage	Untreated sewage	-	(Ng *et al.*, 2012)
JQ898333	Nepavirus	Nepal	Sewage	Untreated sewage	-	(Ng *et al.*, 2012)
JX274036	Porcine associated stool circular virus	New Zealand	Faeces	*Sus scrofa*	Pig	(Sikorski *et al.*, 2013a)
JX305991	Pig stool associated circular ssDNA virus	China	Faeces	*Sus scrofa*	Pig	Unpublished
JX305992	Pig stool associated circular ssDNA virus	China	Faeces	*Sus scrofa*	Pig	Unpublished
JX305993	Pig stool associated circular ssDNA virus	China	Faeces	*Sus scrofa*	Pig	Unpublished
JX305994	Pig stool associated circular ssDNA virus	China	Faeces	*Sus scrofa*	Pig	Unpublished
JX305995	Pig stool associated circular ssDNA virus	China	Faeces	*Sus scrofa*	Pig	Unpublished
JX305996	Pig stool associated circular ssDNA virus	China	Faeces	*Sus scrofa*	Pig	Unpublished
JX305997	Pig stool associated circular ssDNA virus	China	Faeces	*Sus scrofa*	Pig	Unpublished
JX305998	Pig stool associated circular ssDNA virus	China	Faeces	*Sus scrofa*	Pig	Unpublished
JX559621	Circo-like virus-Brazil hs1	Brazil	Faeces	*Homo sapiens*	Human	(Castrignano *et al.*, 2013)
JX559622	Circo-like virus-Brazil hs2	Brazil	Faeces	*Homo sapiens*	Human	(Castrignano *et al.*, 2013)
KC241984	Canine circovirus	USA	Faeces	*Canis lupus familiaris*	Dog	(Li *et al.*, 2013)
KC545226	Porcine stool-associated circular virus 2	USA	Faeces	*Sus scrofa*	Pig	(Cheung *et al.*, 2013)
KC545227	Porcine stool-associated circular virus 3	USA	Faeces	*Sus scrofa*	Pig	(Cheung *et al.*, 2013)
KC545228	Porcine stool-associated circular virus 3	USA	Faeces	*Sus scrofa*	Pig	(Cheung *et al.*, 2013)
KC545229	Porcine stool-associated circular virus 3	USA	Faeces	*Sus scrofa*	Pig	(Cheung *et al.*, 2013)
KC545230	Porcine stool-associated circular virus 3	USA	Faeces	*Sus scrofa*	Pig	(Cheung *et al.*, 2013)
KF031470	Cyclovirus VN	Viet Nam	Faeces	*Sus scrofa*	Pig	(Tan *et al.*, 2013)
KF031471	Cyclovirus VN	Viet Nam	Faeces	*Gallus gallus*	Red junglefowl	(Tan *et al.*, 2013)
KF193403	PoSCV Kor J481	South Korea	Faeces	*Sus scrofa*	Pig	(Kim *et al.*, 2014)
KF246569	Fur seal faeces associated circular DNA virus	New Zealand	Faeces	*Arctocephalus forsteri*	New Zealand fur seal	(Sikorski *et al.*, 2013b)
KF371630	Faecal-associated gemycircularvirus 12	New Zealand	Faeces	*Struthio camelus*	Ostrich	(Sikorski *et al.*, 2013d)
KF371631	Faecal-associated gemycircularvirus 11	New Zealand	Faeces	*Oryctolagus cuniculus*	European rabbit	(Sikorski *et al.*, 2013d)
KF371632	Faecal-associated gemycircularvirus 10	New Zealand	Faeces	*Sturnus vulgaris*	Starling	(Sikorski *et al.*, 2013d)
KF371633	Faecal-associated gemycircularvirus 9	New Zealand	Faeces	*Turdus merula*	Blackbird	(Sikorski *et al.*, 2013d)
KF371634	Faecal-associated gemycircularvirus 8	New Zealand	Faeces	Petroica traversi	Black robin	(Sikorski *et al.*, 2013d)
KF371635	Faecal-associated gemycircularvirus 7	New Zealand	Faeces	*Anas platyrhynchos*	Duck	(Sikorski *et al.*, 2013d)

KF371636	Faecal-associated gemycircularvirus 6	New Zealand	Faeces	*Gerygone albofrontata*	Chatham Island Warbler	(Sikorski *et al.*, 2013d)
KF371637	Faecal-associated gemycircularvirus 5	New Zealand	Faeces	*Gerygone albofrontata*	Chatham Island Warbler	(Sikorski *et al.*, 2013d)
KF371638	Faecal-associated gemycircularvirus 4	New Zealand	Faeces	*Arctocephalus forsteri*	New Zealand Fur Seal	(Sikorski *et al.*, 2013d)
KF371639	Faecal-associated gemycircularvirus 3	New Zealand	Faeces	*Gerygone albofrontata*	Chatham Island Warbler	(Sikorski *et al.*, 2013d)
KF371640	Faecal-associated gemycircularvirus 2	New Zealand	Faeces	*Sus scrofa*	Pig	(Sikorski *et al.*, 2013d)
KF371641	Faecal-associated gemycircularvirus 1c	New Zealand	Faeces	*Turdus merula*	Blackbird	(Sikorski *et al.*, 2013d)
KF371642	Faecal-associated gemycircularvirus 1b	New Zealand	Faeces	*Turdus merula*	Blackbird	(Sikorski *et al.*, 2013d)
KF371643	Faecal-associated gemycircularvirus 1a	New Zealand	Faeces	*Ovis aries*	Sheep	(Sikorski *et al.*, 2013d)
KF880727	Turkey stool associated circular ssDNA virus	Hungary	Faeces	*Meleagris gallopavo*	Domestic turkey	(Reuter *et al.*, 2014)
KJ206566	Human circovirus VS6600022	Netherlands	Faeces	*Homo sapiens*	Human	(Smits *et al.*, 2014)
KJ417134	Porcine stool-associated circular virus 4	USA	Faeces	*Sus scrofa*	Pig	(Cheung *et al.*, 2014b)
KJ433989	Porcine stool-associated circular virus 5	USA	Faeces	*Sus scrofa*	Pig	(Cheung *et al.*, 2014a)
KJ547618	Sewage-associated circular DNA molecule-1	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547619	Sewage-associated circular DNA molecule-3	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547620	Sewage-associated circular DNA virus-1	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547621	Sewage-associated circular DNA virus-10	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547622	Sewage-associated circular DNA virus-11	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547623	Sewage-associated circular DNA virus-12	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547624	Sewage-associated circular DNA virus-13	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547625	Sewage-associated circular DNA virus-14	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547626	Sewage-associated circular DNA virus-2	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547627	Sewage-associated circular DNA virus-3	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547628	Sewage-associated circular DNA virus-4	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547629	Sewage-associated circular DNA virus-5	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547630	Sewage-associated circular DNA virus-6	New Zealand	Sewage	Sewage oxidation Sond	-	(Kraberger *et al.*, 2015a)
KJ547631	Sewage-associated circular DNA virus-7	New Zealand	Sewage	Sewage oxidation Sond	-	(Kraberger *et al.*, 2015a)
KJ547632	Sewage-associated circular DNA virus-8	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547633	Sewage-associated circular DNA virus-9	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547634	Sewage-associated gemycircularvirus-4	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547635	Sewage-associated gemycircularvirus-5	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547636	Sewage-associated gemycircularvirus-6	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547637	Sewage-associated gemycircularvirus-7a	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547638	Sewage-associated gemycircularvirus-8	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547639	Sewage-associated gemycircularvirus-9	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547640	Sewage-associated gemycircularvirus-7b	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547641	Sewage-associated gemycircularvirus-11	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547642	Sewage-associated gemycircularvirus-2	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547643	Sewage-associated gemycircularvirus-3	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547644	Sewage-associated gemycircularvirus-10a	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ547645	Sewage-associated gemycircularvirus-10b	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KJ577810	Porcine stool-associated circular virus-1	USA	Faeces	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577811	Porcine stool-associated circular virus-1	USA	Faeces	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577812	Porcine stool-associated circular virus-7	USA	Faeces	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577813	Porcine stool-associated circular virus-7	USA	Faeces	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577814	Porcine stool-associated circular virus-7	USA	Faeces	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577815	Porcine stool-associated circular virus-7	USA	Faeces	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577816	Porcine stool-associated circular virus-9	USA	Faeces	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577817	Porcine stool-associated circular virus-8	USA	Faeces	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)

KJ577818	Porcine stool-associated circular virus-2	USA	Faeces	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ577819	Porcine stool-associated circular virus-6	USA	Faeces	*Sus scrofa domesticus*	Pig	(Cheung *et al.*, 2015)
KJ641719	Bat gemycircularvirus 23 GD2012	China	Pharyngeal & rectal swabs	*Miniopterus fuliginosus*	Bat	(Wu *et al.*, 2015)
KJ641726	Bat gemycircularvirus 8 NM2013	China	Pharyngeal & rectal swabs	*Rhinolophus ferrumequinum*	Bat	(Wu *et al.*, 2015)
KJ641737	Bat gemycircularvirus Tibet2013	China	Pharyngeal & rectal swabs	*Rhinolophus hipposideros*	Bat	(Wu *et al.*, 2015)
KJ938716	Ancient caribou feces associated virus	Canada	Faeces	*Rangifer tarandus*	Caribou	(Ng *et al.*, 2014)
KJ938717	Caribou feces-associated gemycircularvirus	Canada	Faeces	*Rangifer tarandus*	Caribou	(Ng *et al.*, 2014)
KM017740	Feline cyclovirus	USA	Faeces	*Felis catus*	Cat	(Zhang *et al.*, 2014)
KM382269	Bat circovirus POA/2012/II	Brazil	Faeces	*Molossus molossus, Tadarida brasiliensis*	Bat	(Lima *et al.*, 2015)
KM382270	Bat circovirus POA/2012/VI	Brazil	Faeces	*Molossus molossus, Tadarida brasiliensis*	Bat	(Lima *et al.*, 2015)
KM382271	Bat circovirus POA/2012/I	Brazil	Faeces	*Molossus molossus, Tadarida brasiliensis*	Bat	(Lima *et al.*, 2015)
KM382272	Bat circovirus POA/2012/V	Brazil	Faeces	*Molossus molossus, Tadarida brasiliensis*	Bat	(Lima *et al.*, 2015)
KM392284	Swine cyclovirus	Cameroon	Faeces	*Sus scrofa*	Pig	(Garigliany *et al.*, 2014)
KM392285	Swine cyclovirus	Cameroon	Faeces	*Sus scrofa*	Pig	(Garigliany *et al.*, 2014)
KM392286	Swine cyclovirus	Cameroon	Faeces	*Sus scrofa*	Pig	(Garigliany *et al.*, 2014)
KM392287	Human cyclovirus VN-like	Madagascar	Faeces	*Homo sapiens*	Human	(Garigliany *et al.*, 2014)
KM392288	Human cyclovirus VN-like	Madagascar	Faeces	*Homo sapiens*	Human	(Garigliany *et al.*, 2014)
KM392289	Human cyclovirus VN-like	Madagascar	Faeces	*Homo sapiens*	Human	(Garigliany *et al.*, 2014)
KM573763	Dromedary stool-associated circular ssDNA virus	United Arab Emirates	Faeces	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573764	Dromedary stool-associated circular ssDNA virus	United Arab Emirates	Faeces	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573765	Dromedary stool-associated circular ssDNA virus	United Arab Emirates	Faeces	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573766	Dromedary stool-associated circular ssDNA virus	United Arab Emirates	Faeces	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573767	Dromedary stool-associated circular ssDNA virus	United Arab Emirates	Faeces	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573768	Dromedary stool-associated circular ssDNA virus	United Arab Emirates	Faeces	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573769	Dromedary stool-associated circular ssDNA virus	United Arab Emirates	Faeces	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573770	Dromedary stool-associated circular ssDNA virus	United Arab Emirates	Faeces	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573771	Dromedary stool-associated circular ssDNA virus	United Arab Emirates	Faeces	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573772	Dromedary stool-associated circular ssDNA virus	United Arab Emirates	Faeces	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573773	Dromedary stool-associated circular ssDNA virus	United Arab Emirates	Faeces	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573774	Dromedary stool-associated circular ssDNA virus	United Arab Emirates	Faeces	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573775	Dromedary stool-associated circular ssDNA virus	United Arab Emirates	Faeces	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM573776	Dromedary stool-associated circular ssDNA virus	United Arab Emirates	Faeces	*Camelus dromedarius*	Dromedary camel	(Woo *et al.*, 2014)
KM821747	Sewage-associated gemycircularvirus-1	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821748	Sewage-associated circular DNA virus-36	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821749	Sewage-associated circular DNA virus-37	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821750	Sewage-associated circular DNA virus-15	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821751	Sewage-associated circular DNA virus-16	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821752	Sewage-associated circular DNA virus-17	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821753	Sewage-associated circular DNA virus-18	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821754	Sewage-associated circular DNA virus-19	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821755	Sewage-associated circular DNA virus-20	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821756	Sewage-associated circular DNA virus-21	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821757	Sewage-associated circular DNA virus-22	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821758	Sewage-associated circular DNA virus-23	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821759	Sewage-associated circular DNA virus-24	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821760	Sewage-associated circular DNA virus-25	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821761	Sewage-associated circular DNA virus-26	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821762	Sewage-associated circular DNA virus-27	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)

KM821763	Sewage-associated circular DNA virus-28	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821764	Sewage-associated circular DNA virus-29	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821765	Sewage-associated circular DNA virus-30	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821766	Sewage-associated circular DNA virus-31	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821767	Sewage-associated circular DNA virus-32	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821768	Sewage-associated circular DNA virus-33	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821769	Sewage-associated circular DNA virus-34	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM821770	Sewage-associated circular DNA virus-35	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM877832	Sewage-associated circular DNA molecule-10	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KM877833	Sewage-associated circular DNA molecule-11	New Zealand	Sewage	Sewage oxidation pond	-	(Kraberger *et al.*, 2015a)
KP133078	Gemycircularvirus BZ1	Brazil	Faeces	*Homo sapiens*	Human	(Phan *et al.*, 2015)
KP133079	Gemycircularvirus BZ2	Brazil	Faeces	*Homo sapiens*	Human	(Phan *et al.*, 2015)
KP133080	Gemycircularvirus NP	Nepal	Sewage	Untreated sewage	-	(Phan *et al.*, 2015)
KP151567	Cyclovirus NI-204	Nicaragua	Faeces	*Homo sapiens*	Human	(Phan *et al.*, 2015)
KP233174	Human smacovirus 1	France	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233175	Human smacovirus 1	France	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233176	Human smacovirus 1	France	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233177	Human smacovirus 1	France	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233178	Human smacovirus 1	France	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233179	Human smacovirus 1	France	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233180	Human smacovirus 1	USA	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233181	Human smacovirus 1	USA	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233182	Human smacovirus 1	USA	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233183	Human smacovirus 1	USA	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233184	Human smacovirus 1	USA	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233185	Human smacovirus 1	USA	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233186	Human smacovirus 1	USA	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233187	Human smacovirus 1	USA	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233188	Human smacovirus 1	USA	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233189	Black howler monkey smacovirus	USA	Faeces	*Alouatta caraya*	Black howler monkey	(Ng *et al.*, 2015)
KP233190	Chimpanzee smacovirus	USA	Faeces	*Pan troglodytes*	Chimpanzee	(Ng *et al.*, 2015)
KP233191	Gorilla smacovirus	USA	Faeces	*Gorilla gorilla*	Gorilla	(Ng *et al.*, 2015)
KP233192	Gorilla smacovirus	USA	Faeces	*Gorilla gorilla*	Gorilla	(Ng *et al.*, 2015)
KP233193	Human smacovirus 1	USA	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP233194	Lemur smacovirus	USA	Faeces	*Lemur catta*	Ring-tailed lemur	(Ng *et al.*, 2015)
KP263543	Badger feces-associated gemycircularvirus	Portugal	Faeces	*Meles meles*	European Badger	(Conceicao-Neto *et al.*, 2015)
KP263544	Mongoose feces-associated gemycircularvirus a	Portugal	Faeces	*Herpestes ichneumon*	Egyptian Mongoose	(Conceicao-Neto *et al.*, 2015)
KP263545	Mongoose feces-associated gemycircularvirus b	Portugal	Faeces	*Herpestes ichneumon*	Egyptian Mongoose	(Conceicao-Neto *et al.*, 2015)
KP263546	Mongoose feces-associated gemycircularvirus c	Portugal	Faeces	*Herpestes ichneumon*	Egyptian Mongoose	(Conceicao-Neto *et al.*, 2015)
KP263547	Mongoose feces-associated gemycircularvirus d	Portugal	Faeces	*Herpestes ichneumon*	Egyptian Mongoose	(Conceicao-Neto *et al.*, 2015)
KP264964	Human smacovirus 1	France	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP264965	Human smacovirus 1	France	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP264966	Human smacovirus 1	France	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP264967	Human smacovirus 1	France	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP264968	Human smacovirus 1	France	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP264969	Human smacovirus 1	France	Faeces	*Homo sapiens*	Human	(Ng *et al.*, 2015)
KP860906	Rat stool-associated circular ssDNA virus	Germany	Faeces; intestinal content	*Rattus norvegicus*	Rat	(Sachsenröder *et al.*, 2014)
KP860907	Rat stool-associated circular ssDNA virus	Germany	Faeces; intestinal content	*Rattus norvegicus*	Rat	(Sachsenröder *et al.*, 2014)

KP860908	Rat stool-associated circular ssDNA virus	Germany	Faeces; intestinal content	*Rattus norvegicus*	Rat	(Sachsenröder *et al.*, 2014)
KP941114	Fox circovirus	Croatia	Faeces	*Vulpes vulpes*	Red Fox	Unpublished
KT600065	Human feces PeCV-PE	Peru	Faeces	*Homo sapiens*	Human	(Phan *et al.*, 2016)
KT600066	Human feces PeCV-NI	Nicaragua	Faeces	*Homo sapiens*	Human	(Phan *et al.*, 2016)
KT600067	Human feces PeCV-CH	Chile	Faeces	*Homo sapiens*	Human	(Phan *et al.*, 2016)
KT600068	Human feces smacovirus 2	Peru	Faeces	*Homo sapiens*	Human	(Phan *et al.*, 2016)
KT600069	Human feces smacovirus 3	Peru	Faeces	*Homo sapiens*	Human	(Phan *et al.*, 2016)
KT732783	Pacific flying fox faeces associated circular DNA molecule-1	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732784	Pacific flying fox faeces associated circular DNA virus-9	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732785	Pacific flying fox associated cyclovirus-1	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732786	Pacific flying fox associated cyclovirus-2	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732787	Pacific flying fox associated cyclovirus-3	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732788	Pacific flying fox associated cyclovirus-3	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732789	Pacific flying fox associated cyclovirus-3	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732790	Pacific flying fox faeces associated gemycircularvirus-1	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732791	Pacific flying fox faeces associated gemycircularvirus-1	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732792	Pacific flying fox faeces associated gemycircularvirus-2	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732793	Pacific flying fox faeces associated gemycircularvirus-2	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732794	Pacific flying fox faeces associated gemycircularvirus-3	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732795	Pacific flying fox faeces associated gemycircularvirus-4	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732796	Pacific flying fox faeces associated gemycircularvirus-4	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732797	Pacific flying fox faeces associated gemycircularvirus-5	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732798	Pacific flying fox faeces associated gemycircularvirus-6	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732799	Pacific flying fox faeces associated gemycircularvirus-6	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732800	Pacific flying fox faeces associated gemycircularvirus-7	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732801	Pacific flying fox faeces associated gemycircularvirus-8	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732802	Pacific flying fox faeces associated gemycircularvirus-8	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732803	Pacific flying fox faeces associated gemycircularvirus-9	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732804	Pacific flying fox faeces associated gemycircularvirus-10	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732805	Pacific flying fox faeces associated gemycircularvirus-10	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732806	Pacific flying fox faeces associated gemycircularvirus-14	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732807	Pacific flying fox faeces associated gemycircularvirus-11	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732808	Pacific flying fox faeces associated gemycircularvirus-11	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732809	Pacific flying fox faeces associated gemycircularvirus-11	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732810	Pacific flying fox faeces associated gemycircularvirus-11	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732811	Pacific flying fox faeces associated gemycircularvirus-11	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732812	Pacific flying fox faeces associated gemycircularvirus-11	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732813	Pacific flying fox faeces associated gemycircularvirus-12	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732814	Pacific flying fox faeces associated gemycircularvirus-13	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732815	Pacific flying fox associated multicomponent virus-1	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732816	Pacific flying fox associated multicomponent virus-1	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732817	Pacific flying fox associated multicomponent virus-1	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732818	Pacific flying fox faeces associated circular DNA virus-3	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732819	Pacific flying fox faeces associated circular DNA virus-4	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732820	Pacific flying fox faeces associated circular DNA virus-1	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732821	Pacific flying fox faeces associated circular DNA virus-1	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732822	Pacific flying fox faeces associated circular DNA virus-5	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732823	Pacific flying fox faeces associated circular DNA virus-6	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)

KT732824	Pacific flying fox faeces associated circular DNA virus-7	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732825	Pacific flying fox faeces associated circular DNA virus-8	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732826	Pacific flying fox faeces associated circular DNA molecule-2	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732827	Pacific flying fox faeces associated circular DNA virus-10	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732828	Pacific flying fox faeces associated circular DNA virus-11	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732829	Pacific flying fox faeces associated circular DNA virus-2	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732830	Pacific flying fox faeces associated circular DNA virus-12	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732831	Pacific flying fox faeces associated circular DNA virus-2	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732832	Pacific flying fox faeces associated circular DNA virus-13	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732833	Pacific flying fox faeces associated circular DNA virus-14	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT732834	Pacific flying fox faeces associated circular DNA virus-15	Tonga	Faeces	*Pteropus tonganus*	Pacific flying fox	(Male *et al.*, 2016)
KT945154	Rat stool-associated circular ssDNA virus	Denmark	Faeces	*Rattus norvegicus*	Rat	(Hansen *et al.*, 2015)
LC018134	Taiwan squirrel cyclovirus-1	Japan	Intestinal contents	Callosciurus erythraeus thaiwanensis	Taiwan squirrels	(Sato *et al.*, 2015)

References

Abouzid, A. M., Frischmuth, T. & Jeske, H. (1988). A putative replicative form of the abutilon mosaic virus (gemini group) in a chromatin-like structure. *MGG Molecular & General Genetics* **212**, 252-258.

Abrahão, J. S., Trindade, G. S., Ferreira, J. M., Campos, R. K., Bonjardim, C. A., Ferreira, P. C. & Kroon, E. G. (2009). Long-lasting stability of Vaccinia virus strains in murine feces: implications for virus circulation and environmental maintenance. *Archives of Virology* **154**, 1551-1553.

Adriaenssens, E. M., Van Zyl, L., De Maayer, P., Rubagotti, E., Rybicki, E., Tuffin, M. & Cowan, D. A. (2015). Metagenomic analysis of the viral community in Namib Desert hypoliths. *Environmental Microbiology* **17**, 480-495.

Alberter, B., Ali Rezaian, M. & Jeske, H. (2005). Replicative intermediates of Tomato leaf curl virus and its satellite DNAs. *Virology* **331**, 441-448.

Allan, G. M., McNeilly, F., Kennedy, S., Daft, B., Clarke, E. G., Ellis, J. A., Haines, D. M., Meehan, B. M. & Adair, B. M. (1998). Isolation of porcine circovirus-like viruses from pigs with a wasting disease in the USA and Europe. *Journal of Veterinary Diagnostic Investigation* **10**, 3-10.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Ambert-Balay, K. & Pothier, P. (2013). Evaluation of 4 immunochromatographic tests for rapid detection of norovirus in faecal samples. *Journal of Clinical Virology* **56**, 194-198.

Amin, I., Mansoor, S., Amrao, L., Hussain, M., Irum, S., Zafar, Y., Bull, S. E. & Briddon, R. W. (2006). Mobilisation into cotton and spread of a recombinant cotton leaf curl disease satellite. *Handbook of Environmental Chemistry, Volume 5: Water Pollution* **151**, 2055-2065.

Angly, F. E., Felts, B., Breitbart, M., Salamon, P., Edwards, R. A., Carlson, C., Chan, A. M., Haynes, M., Kelley, S., Liu, H., Mahaffy, J. M., Mueller, J. E., Nulton, J., Olson, R., Parsons, R., Rayhawk, S., Suttle, C. A. & Rohwer, F. (2006). The marine viromes of four oceanic regions. *PLoS Biology* **4**, e368.

Argüello-Astorga, G. R. & Ruiz-Medrano, R. (2001). An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: Identification of potential interacting amino acid-base pairs by a comparative approach. *Archives of Virology* **146**, 1465-1485.

Atanasova, N. S., Roine, E., Oren, A., Bamford, D. H. & Oksanen, H. M. (2012). Global network of specific virus-host interactions in hypersaline environments. *Environmental Microbiology* **14**, 426-440.

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological reviews* **35**, 235-241.

Barreto, S. C., Uppalapati, M. & Ray, A. (2014). Small circular DNAs in human pathology. *Malaysian Journal of Medical Sciences* **21**, 4-18.

Basso, M. F., da Silva, J. C., Fajardo, T. V., Fontes, E. P. & Zerbini, F. M. (2015). A novel, highly divergent ssDNA virus identified in Brazil infecting apple, pear and grapevine. *Virus Research* **210**, 27-33.

Bavelaar, H. H. J., Rahamat-Langendoen, J., Niesters, H. G. M., Zoll, J. & Melchers, W. J. G. (2015). Whole genome sequencing of fecal samples as a tool for the diagnosis and genetic characterization of norovirus. *Journal of Clinical Virology* **72**, 122-125.

Bille, E., Zahar, J. R., Perrin, A., Morelle, S., Kriz, P., Jolley, K. A., Maiden, M. C. J., Dervin, C., Nassif, X. & Tinsley, C. R. (2005). A chromosomally integrated bacteriophage in invasive meningococci. *Journal of Experimental Medicine* **201**, 1905-1913.

Bize, A., Peng, X., Prokofeva, M., MacLellan, K., Lucas, S., Forterre, P., Garrett, R. A., Bonch-Osmolovskaya, E. A. & Prangishvili, D. (2008). Viruses in acidic geothermal environments of the Kamchatka Peninsula. *Research in Microbiology* **159**, 358-366.

Blanco, L., Bernad, A., Lazaro, J. M., Martin, G., Garmendia, C. & Salas, M. (1989). Highly efficient DNA synthesis by the phage Φ29 DNA polymerase. Symmetrical mode of DNA replication. *Journal of Biological Chemistry* **264**, 8935-8940.

Blinkova, O., Victoria, J., Li, Y., Keele, B. F., Sanz, C., Ndjango, J. B. N., Peeters, M., Travis, D., Lonsdorf, E. V., Wilson, M. L., Pusey, A. E., Hahn, B. H. & Delwart, E. L. (2010). Novel circular DNA viruses in stool samples of wild-living chimpanzees. *Journal of General Virology* **91**, 74-86.

Bolduc, B., Shaughnessy, D. P., Wolf, Y. I., Koonin, E. V., Roberto, F. F. & Young, M. (2012). Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated yellowstone hot springs. *Journal of Virology* **86**, 5562-5573.

Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S. B., Krstic, P. S., Lindsay, S., Ling, X. S., Mastrangelo, C. H., Meller, A., Oliver, J. S., Pershin, Y. V., Ramsey, J. M., Riehn, R., Soni, G. V., Tabard-Cossa, V., Wanunu, M., Wiggin, M. & Schloss, J. A. (2008). The potential and challenges of nanopore sequencing. *Nature Biotechnology* **26**, 1146-1153.

Breitbart, M., Benner, B. E., Jernigan, P. E., Rosario, K., Birsa, L. M., Harbeitner, R., Fulford, S., Graham, C., Walters, A., Goldsmith, D. B., Berger, S. A. & Nejstgaard, J. C. (2015). Discovery, prevalence, and persistence of novel circular single-stranded DNA viruses in the ctenophores Mnemiopsis leidyi and Beroe ovata. *Frontiers in Microbiology* **6**, 1426.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P. & Rohwer, F. (2003). Metagenomic analyses of an uncultured viral community from human feces. *Journal of Bacteriology* **185**, 6220-6223.

Breitbart, M., Wegley, L., Leeds, S., Schoenfeld, T. & Rohwer, F. (2004). Phage Community Dynamics in Hot Springs. *Applied and Environmental Microbiology* **70**, 1633-1640.

Briddon, R. W., Bull, S. E., Amin, I., Idris, A. M., Mansoor, S., Bedford, I. D., Dhawan, P., Rishi, N., Siwatch, S. S., Abdel-Salam, A. M., Brown, J. K., Zafar, Y. & Markham, P. G. (2003). Diversity of DNA β, a

satellite molecule associated with some monopartite begomoviruses. *Virology* **312**, 106-121.

Briddon, R. W., Patil, B. L., Bagewadi, B., Nawaz-Ul-Rehman, M. S. & Fauquet, C. M. (2010). Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. *BMC Evolutionary Biology* **10**.

Brough, C. L., Gardiner, W. E., Inamdar, N. M., Zhang, X. Y., Ehrlich, M. & Bisaro, D. M. (1992). DNA methylation inhibits propagation of tomato golden mosaic virus DNA in transfected protoplasts. *Plant Molecular Biology* **18**, 703-712.

Brown, J. K., Idris, A. M., Alteri, C. & Stenger, D. C. (2002). Emergence of a new Cucurbit-infecting begomovirus species capable of forming viable reassortants with related viruses in the Squash leaf curl virus cluster. *Phytopathology* **92**, 734-742.

Burns, T. M., Harding, R. M. & Dale, J. L. (1995). The genome organization of banana bunchy top virus: Analysis of six ssDNA components. *Journal of General Virology* **76**, 1471-1482.

Cadar, D., Kiss, T., Ádám, D., Cságola, A., Novosel, D. & Tuboly, T. (2013). Phylogeny, spatio-temporal phylodynamics and evolutionary scenario of Torque teno sus virus 1 (TTSuV1) and 2 (TTSuV2) in wild boars: Fast dispersal and high genetic diversity. *Veterinary Microbiology* **166**, 200-213.

Cai, L., Ni, J., Xia, Y., Zi, Z., Ning, K., Qiu, P., Li, X., Wang, B., Liu, Q., Hu, D., Yu, X., Zhou, Z., Zhai, X., Han, X. & Tian, K. (2012). Identification of an emerging recombinant cluster in porcine circovirus type 2. *Virus Research* **165**, 95-102.

Campos, J., Martínez, E., Izquierdo, Y. & Fando, R. (2010). VEJφ, a novel filamentous phage of Vibrio cholerae able to transduce the cholera toxin genes. *Microbiology* **156**, 108-115.

Cantalupo, P. G., Calgua, B., Zhao, G., Hundesa, A., Wier, A. D., Katz, J. P., Grabe, M., Hendrix, R. W., Girones, R., Wang, D. & Pipas, J. M. (2011). Raw sewage harbors diverse viral populations. *mBio* **2**.

Castrignano, S. B., Nagasse-Sugahara, T. K., Kisielius, J. J., Ueda-Ito, M., Brandão, P. E. & Curti, S. P. (2013). Two novel circo-like viruses detected in human feces: Complete genome sequencing and electron microscopy analysis. *Virus Research* **178**, 364-373.

Chakraborty, S., Vanitharani, R., Chattopadhyay, B. & Fauquet, C. M. (2008). Supervirulent pseudorecombination and asymmetric synergism between genomic components of two distinct species of begomovirus associated with severe tomato leaf curl disease in India. *Journal of General Virology* **89**, 818-828.

Cheung, A. K. (2004). Identification of an octanucleotide motif sequence essential for viral protein, DNA, and progeny virus biosynthesis at the origin of DNA replication of porcine circovirus type 2. *Virology* **324**, 28-36.

Cheung, A. K. (2009). Homologous recombination within the capsid gene of porcine circovirus type 2 subgroup viruses via natural co-infection. *Archives of Virology* **154**, 531-534.

Cheung, A. K. (2012). Porcine circovirus: Transcription and DNA replication. *Virus Research* **164**, 46-53.

Cheung, A. K., Ng, T. F., Lager, K. M., Alt, D. P., Delwart, E. L. & Pogranichniy, R. M. (2014a). Identification of a novel single-stranded circular DNA virus in pig feces. *Genome announcements* **2**, 00347-00314.

Cheung, A. K., Ng, T. F., Lager, K. M., Alt, D. P., Delwart, E. L. & Pogranichniy, R. M. (2014b). Unique circovirus-like genome detected in pig feces. *Genome announcements* **2**, 00251-00214.

Cheung, A. K., Ng, T. F., Lager, K. M., Bayles, D. O., Alt, D. P., Delwart, E. L., Pogranichniy, R. M. & Kehrli, M. E. (2013). A divergent clade of circular single-stranded DNA viruses from pig feces. *Archives of Virology* **158**, 2157-2162.

Cheung, A. K., Ng, T. F. F., Lager, K. M., Alt, D. P., Delwart, E. & Pogranichniy, R. M. (2015). Identification of several clades of novel single-stranded circular DNA viruses with conserved stem-loop structures in pig feces. *Archives of Virology* **160**, 353-358.

Choudhury, N. R., Malik, P. S., Singh, D. K., Islam, M. N., Kaliappan, K. & Mukherjee, S. K. (2006). The oligomeric Rep protein of Mungbean yellow mosaic India virus (MYMIV) is a likely replicative helicase. *Nucleic Acids Research* **34**, 6362-6377.

Clérot, D. & Bernardi, F. (2006). DNA helicase activity is associated with the replication initiator protein Rep of tomato yellow leaf curl germinivirus. *Journal of Virology* **80**, 11322-11330.

Collin, S., Fernández-Lobato, M., Gooding, P. S., Mullineaux, P. M. & Fenoll, C. (1996). The Two Nonstructural Proteins from Wheat Dwarf Virus Involved in Viral Gene Expression and Replication Are Retinoblastoma-Binding Proteins. *Virology* **219**, 324-329.

Conceicao-Neto, N., Zeller, M., Heylen, E., Lefrere, H., Mesquita, J. R. & Matthijnssens, J. (2015). Fecal virome analysis of three carnivores reveals a novel nodavirus and multiple gemycircularviruses. *Virology Journal* **12**, e79.

Cotmore, S. F., Agbandje-McKenna, M., Chiorini, J. A., Mukha, D. V., Pintel, D. J., Qiu, J., Soderlund-Venermo, M., Tattersall, P., Tijssen, P., Gatherer, D. & Davison, A. J. (2014). The family Parvoviridae. *Archives of Virology* **159**, 1239-1247.

Crane, P. E., Hopkins, A. J. M., Dick, M. A. & Bulman, L. S. (2009). Behaviour of Neonectria fuckeliana causing a pine canker disease in New Zealand. *Canadian Journal of Forest Research* **39**, 2119-2128.

Cromie, G. A., Connelly, J. C. & Leach, D. R. F. (2001). Recombination at double-strand breaks and DNA ends: Conserved mechanisms from phage to humans. *Molecular Cell* **8**, 1163-1174.

Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164-1165.

Davino, S., Napoli, C., Dellacroce, C., Miozzi, L., Noris, E., Davino, M. & Accotto, G. P. (2009). Two new natural begomovirus recombinants associated with the tomato yellow leaf curl disease co-exist with parental viruses in tomato epidemics in Italy. *Virus Research* **143**, 15-23.

Dayaram, A., Galatowitsch, M., Harding, J. S., Arguello-Astorga, G. R. & Varsani, A. (2014). Novel circular DNA viruses identified in Procordulia grayi and Xanthocnemis zealandica larvae using metagenomic approaches.

Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases **22**, 134-141.

Dayaram, A., Galatowitsch, M. L., Arguello-Astorga, G. R., van Bysterveldt, K., Kraberger, S., Stainton, D., Harding, J. S., Roumagnac, P., Martin, D. P., Lefeuvre, P. & Varsani, A. (2016). Diverse circular replication-associated protein encoding viruses circulating in invertebrates within a lake ecosystem. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*.

Dayaram, A., Goldstien, S., Arguello-Astorga, G. R., Zawar-Reza, P., Gomez, C., Harding, J. S. & Varsani, A. (2015a). Diverse small circular DNA viruses circulating amongst estuarine molluscs. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **31**, 284-295.

Dayaram, A., Goldstien, S., Zawar-Reza, P., Gomez, C., Harding, J. S. & Varsani, A. (2013a). Identification of Starling Circovirus in an Estuarine Mollusc (Amphibola crenata) in New Zealand Using Metagenomic Approaches. *Genome announcements* **1**, e00278-00213.

Dayaram, A., Goldstien, S., Zawar-Reza, P., Gomez, C., Harding, J. S. & Varsani, A. (2013b). Novel ssDNA virus recovered from estuarine mollusc (Amphibola crenata) whose replication associated protein (Rep) shares similarities with Rep-like sequences of bacterial origin. *Journal of General Virology* **94**, 1104-1110.

Dayaram, A., Opong, A., Jäschke, A., Hadfield, J., Baschiera, M., Dobson, R. C. J., Offei, S. K., Shepherd, D. N., Martin, D. P. & Varsani, A. (2012). Molecular characterisation of a novel cassava associated circular ssDNA virus. *Virus Research* **166**, 130-135.

Dayaram, A., Potter, K. A., Moline, A. B., Rosenstein, D. D., Marinov, M., Thomas, J. E., Breitbart, M., Rosario, K., Argüello-Astorga, G. R. & Varsani, A. (2013c). High global diversity of cycloviruses amongst dragonflies. *Journal of General Virology* **94**, 1827-1840.

Dayaram, A., Potter, K. A., Pailes, R., Marinov, M., Rosenstein, D. D. & Varsani, A. (2015b). Identification of diverse circular single-stranded DNA viruses in adult dragonflies and damselflies (Insecta: Odonata) of Arizona and Oklahoma, USA. *Infection, Genetics and Evolution* **30**, 278-287.

de Carvalho Ferreira, H. C., Weesendorp, E., Quak, S., Stegeman, J. A. & Loeffen, W. L. A. (2014). Suitability of faeces and tissue samples as a basis for non-invasive sampling for African swine fever in wild boar. *Veterinary Microbiology* **172**, 449-454.

Dekker, E. L., Woolston, C. J., Xue, Y. B., Cox, B. & Mullineaux, P. M. (1991). Transcript mapping reveals different expression strategies for the bicistronic RNAs of the geminivirus wheat dwarf virus. *Nucleic Acids Research* **19**, 4075-4081.

Delwart, E. & Li, L. (2012). Rapidly expanding genetic diversity and host range of the Circoviridae viral family and other Rep encoding small circular ssDNA genomes. *Virus Research* **164**, 114-121.

Desbiez, C., David, C., Mettouchi, A., Laufs, J. & Gronenborn, B. (1995). Rep protein of tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 5640-5644.

Diemer, G. S. & Stedman, K. M. (2012). A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biology Direct* **7**.

Du, Z., Tang, Y., Zhang, S., She, X., Lan, G., Varsani, A. & He, Z. (2014). Identification and molecular characterization of a single-stranded circular DNA virus with similarities to Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1. *Archives of Virology* **159**, 1527-1531.

Duffy, S. & Holmes, E. C. (2008). Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus Tomato yellow leaf curl virus. *Journal of Virology* **82**, 957-965.

Duffy, S. & Holmes, E. C. (2009). Validation of high rates of nucleotide substitution in geminiviruses: Phylogenetic evidence from East African cassava mosaic viruses. *Journal of General Virology* **90**, 1539-1547.

Dunlap, D. S., Ng, T. F. F., Rosario, K., Barbosa, J. G., Greco, A. M., Breitbart, M. & Hewson, I. (2013). Molecular and microscopic evidence of viruses in marine copepods. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1375-1380.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., DeWinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J. & Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133-138.

Ellis, J., Hassard, L., Clark, E., Harding, J., Allan, G., Willson, P., Strokappe, J., Martin, K., McNeilly, F., Meehan, B., Todd, D. & Haines, D. (1998). Isolation of circovirus from lesions of pigs with postweaning multisystemic wasting syndrome. *Canadian Veterinary Journal* **39**, 44-51.

Emerson, J. B., Thomas, B. C., Andrade, K., Allen, E. E., Heidelberg, K. B. & Banfielda, J. F. (2012). Dynamic viral populations in hypersaline systems as revealed by metagenomic assembly. *Applied and Environmental Microbiology* **78**, 6309-6320.

Fahsbender, E., Hewson, I., Rosario, K., Tuttle, A. D., Varsani, A. & Breitbart, M. (2015). Discovery of a novel circular DNA virus in the Forbes sea star, Asterias forbesi. *Archives of Virology*.

Fancello, L., Trape, S., Robert, C., Boyer, M., Popgeorgiev, N., Raoult, D. & Desnues, C. (2013). Viruses in the desert: A metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. *ISME Journal* **7**, 359-369.

Frischmuth, T. & Stanley, J. (1998). Recombination between viral DNA and the transgenic coat protein gene of African cassava mosaic geminivirus. *Journal of General Virology* **79**, 1265-1271.

Garigliany, M. M., Boerstler, J., Joest, H., Badusche, M., Desmecht, D., Schmidt-Chanasit, J. & Cadar, D. (2015). Characterization of a novel circo-like virus in Aedes vexans mosquitoes from Germany: evidence for a

new genus within the family Circoviridae. *Journal of General Virology* **96**, 915-920.

Garigliany, M. M., Hagen, R. M., Frickmann, H., May, J., Schwarz, N. G., Perse, A., Jöst, H., Börstler, J., Shahhosseini, N., Desmecht, D., Mbunkah, H. A., Daniel, A. M., Kingsley, M. T., De Mendonca Campos, R., De Paula, V. S., Randriamampionona, N., Poppert, S., Tannich, E., Rakotozandrindrainy, R., Cadar, D. & Schmidt-Chanasit, J. (2014). Cyclovirus CyCV-VN species distribution is not limited to Vietnam and extends to Africa. *Scientific Reports* **4**.

Garkavenko, O., Elliott, R. B. & Croxson, M. C. (2005). Identification of pig circovirus type 2 in New Zealand pigs. *Transplantation proceedings* **37**, 506-509.

Garmendia, C., Bernad, A., Esteban, J. A., Blanco, L. & Salas, M. (1992). The bacteriophage φ29 DNA polymerase, a proofreading enzyme. *Journal of Biological Chemistry* **267**, 2594-2599.

Ge, L., Zhang, J., Zhou, X. & Li, H. (2007). Genetic structure and population variability of Tomato yellow leaf curl China virus. *Journal of Virology* **81**, 5902-5907.

Ge, X., Li, J., Peng, C., Wu, L., Yang, X., Wu, Y., Zhang, Y. & Shi, Z. (2011). Genetic diversity of novel circular ssDNA viruses in bats in China. *Journal of General Virology* **92**, 2646-2653.

Ge, X., Li, Y., Yang, X., Zhang, H., Zhou, P., Zhang, Y. & Shi, Z. (2012). Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *Journal of Virology* **86**, 4620-4630.

George, B., Ruhel, R., Mazumder, M., Sharma, V. K., Jain, S. K., Gourinath, S. & Chakraborty, S. (2014). Mutational analysis of the helicase domain of a replication initiator protein reveals critical roles of Lys 272 of the B' motif and Lys 289 of the β-hairpin loop in geminivirus replication. *Journal of General Virology* **95**, 1591-1602.

Geslin, C., Le Romancer, M., Gaillard, M., Erauso, G. & Prieur, D. (2003). Observation of virus-like particles in high temperature enrichment cultures from deep-sea hydrothermal vents. *Research in Microbiology* **154**, 303-307.

Gilbert, W. & Dressler, D. (1968). DNA replication: the rolling circle model. *Cold Spring Harbor Symposia on Quantitative Biology* **33**, 473-484.

Gorbalenya, A. E., Koonin, E. V. & Wolf, Y. I. (1990). A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS Letters* **262**, 145-148.

Greninger, A. L. & DeRisi, J. L. (2015). Draft Genome Sequences of Marine RNA Viruses SF-1, SF-2, and SF-3 Recovered from San Francisco Wastewater. *Genome Announcements* **3**.

Grigoras, I., Ginzo, A. I. C., Martin, D. P., Varsani, A., Romero, J., Mammadov, A. C., Huseynova, I. M., Aliyev, J. A., Kheyr-Pour, A., Huss, H., Ziebell, H., Timchenko, T., Vetten, H. J. & Gronenborn, B. (2014). Genome diversity and evidence of recombination and reassortment in nanoviruses from Europe. *Journal of General Virology* **95**, 1178-1191.

Grigoras, I., Timchenko, T., Grande-Pérez, A., Katul, L., Vetten, H. J. & Gronenborn, B. (2010). High variability and rapid evolution of a nanovirus. *Journal of Virology* **84**, 9105-9117.

Grigoras, I., Timchenko, T., Katul, L., Grande-Pérez, A., Vetten, H. J. & Gronenborn, B. (2009). Reconstitution of authentic nanovirus from multiple cloned DNAs. *Journal of Virology* **83**, 10778-10787.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**, 307-321.

Gutierrez, C. (1999). Geminivirus DNA replication. *Cellular and Molecular Life Sciences* **56**, 313-329.

Ha, H. J., Alley, M. R., Cahill, J. I., Howe, L. & Gartrell, B. D. (2009). The prevalence of psittacine beak and feather disease virus infection in native parrots in New Zealand. *New Zealand Veterinary Journal* **57**, 50-52.

Ha, H. J., Anderson, I. L., Alley, M. R., Springett, B. P. & Gartrell, B. D. (2007). The prevalence of beak and feather disease virus infection in wild populations of parrots and cockatoos in New Zealand. *New Zealand Veterinary Journal* **55**, 235-238.

Hafner, G. J., Stafford, M. R., Wolter, L. C., Harding, R. M. & Dale, J. L. (1997). Nicking and joining activity of banana bunchy top virus replication protein in vitro. *Journal of General Virology* **78**, 1795-1799.

Hanna, Z. R., Runckel, C., Fuchs, J., DeRisi, J. L., Mindell, D. P., Van Hemert, C., Handel, C. M. & Dumbacher, J. P. (2015). Isolation of a complete circular virus genome sequence from an Alaskan black-capped chickadee (Poecile atricapillus) gastrointestinal tract sample. *Genome Announcements* **3**, 01081-01015.

Hansen, T. A., Fridholm, H., Froslev, T. G., Kjartansdottir, K. R., Willerslev, E., Nielsen, L. P. & Hansen, A. J. (2015). New Type of Papillomavirus and Novel Circular Single Stranded DNA Virus Discovered in Urban Rattus norvegicus Using Circular DNA Enrichment and Metagenomics. *PLoS ONE* **10**, e0141952.

Harding, R. M., Burns, T. M. & Dale, J. L. (1991). Virus-like particles associated with banana bunchy top disease contain small single-stranded DNA. *Journal of General Virology* **72**, 225-230.

Häring, M., Rachel, R., Peng, X., Garrett, R. A. & Prangishvili, D. (2005). Viral diversity in hot springs of Pozzuoli, Italy, and characterization of a unique archaeal virus, Acidianus bottle-shaped virus, from a new family, the Ampullaviridae. *Journal of Virology* **79**, 9904-9911.

Harkins, G. W., Delport, W., Duffy, S., Wood, N., Monjane, A. L., Owor, B. E., Donaldson, L., Saumtally, S., Triton, G., Briddon, R. W., Shepherd, D. N., Rybicki, E. P., Martin, D. P. & Varsani, A. (2009). Experimental evidence indicating that mastreviruses probably did not co-diverge with their hosts. *Virology Journal* **6**.

Helmrich, A., Ballarino, M., Nudler, E. & Tora, L. (2013). Transcription-replication encounters, consequences and genomic instability. *Nature Structural and Molecular Biology* **20**, 412-418.

Hendrix, R. W., Lawrence, J. G., Hatfull, G. F. & Casjens, S. (2000). The origins and ongoing evolution of viruses. *Trends in Microbiology* **8**, 504-508.

Hewson, I., Eaglesham, J. B., Hook, T. O., LaBarre, B. A., Sepulveda, M. S., Thompson, P. D., Watkins, J. M. & Rudstam, L. G. (2013a). Investigation of viruses in Diporeia spp. from the Laurentian Great Lakes

and Owasco Lake as potential stressors of declining populations. *Journal of Great Lakes Research* **39**, 499-506.

Hewson, I., Ng, G., Li, W. F., LaBarre, B. A., Aguirre, I., Barbosa, J. G., Breitbart, M., Greco, A. W., Kearns, C. M., Looi, A., Schaffner, L. R., Thompson, P. D. & Hairston, N. G. (2013b). Metagenomic identification, seasonal dynamics, and potential transmission mechanisms of a Daphnia-associated single-stranded DNA virus in two temperate lakes. *Limnology and Oceanography* **58**, 1605-1620.

Heyraud-Nitschke, F., Schumacher, S., Laufs, J., Schaefer, S., Schell, J. & Gronenborn, B. (1995). Determination of the origin cleavage and joining domain of geminivirus Rep proteins. *Nucleic Acids Research* **23**, 910-916.

Hill, J. E., Strandberg, J. O., Hiebert, E. & Lazarowitz, S. G. (1998). Asymmetric infectivity of pseudorecombinants of cabbage leaf curl virus and squash leaf curl virus: Implications for bipartite geminivirus evolution and movement. *Virology* **250**, 283-292.

Hogenhout, S. A., Ammar, E. D., Whitfield, A. E. & Redinbaugh, M. G. (2008). Insect vector interactions with persistently transmitted viruses. In *Annual Review of Phytopathology*, pp. 327-359.

Hood, I. A., Oliva, J., Kimberley, M. O., Arhipova, N. & Bakys, R. (2015). Armillaria novae-zelandiae and other basidiomycete wood decay fungi in New Zealand Pinus radiata thinning stumps. *Forest Pathology* **45**, 298-310.

Horser, C. L., Harding, R. M. & Dale, J. L. (2001a). Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein. *Journal of General Virology* **82**, 459-464.

Horser, C. L., Karan, M., Harding, R. M. & Dale, J. L. (2001b). Additional Rep-encoding DNAs associated with banana bunchy top virus. *Archives of Virology* **146**, 71-86.

Hou, Y. M. & Gilbertson, R. L. (1996). Increased pathogenicity in a pseudorecombinant bipartite geminivirus correlates with intermolecular recombination. *Journal of Virology* **70**, 5430-5436.

Hsia, R. C., Ting, L. M. & Bavoil, P. M. (2000). Microvirus of Chlamydia psittaci strain Guinea pig inclusion conjunctivitis: Isolation and molecular characterization. *Microbiology* **146**, 1651-1660.

Hu, J. M., Fu, H. C., Lin, C. H., Su, H. J. & Yeh, H. H. (2007). Reassortment and concerted evolution in Banana bunchy top virus genomes. *Journal of Virology* **81**, 1746-1761.

Hu, Z. Y., Li, G. H., Li, G. T., Yao, Q. & Chen, K. P. (2013). Bombyx mori bidensovirus: The type species of the new genus Bidensovirus in the new family Bidnaviridae. *Chinese Science Bulletin* **58**, 4528-4532.

Huang, C., Xie, Y., Zhao, L., Ren, H. & Li, Z. (2013). A naturally occurring defective DNA satellite associated with a monopartite begomovirus: Evidence for recombination between alphasatellite and betasatellite. *Viruses* **5**, 2116-2128.

Hughes, A. L. (2004). Birth-and-death evolution of protein-coding regions and concerted evolution of non-coding regions in the multi-component genomes of nanoviruses. *Molecular Phylogenetics and Evolution* **30**, 287-294.

Idris, A. M. & Brown, J. K. (2004). Cotton leaf crumple virus is a distinct Western Hemisphere begomovirus species with complex evolutionary

relationships indicative of recombination and reassortment. *Phytopathology* **94**, 1068-1074.

Ilyina, T. V. & Koonin, E. V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Research* **20**, 3279-3285.

Inamdar, N. M., Zhang, X. Y., Brough, C. L., Gardiner, W. E., Bisaro, D. M. & Ehrlich, M. (1992). Transfection of heteroduplexes containing uracil · guanine or thymine · guanine mispairs into plant cells. *Plant Molecular Biology* **20**, 123-131.

Itoh, Y., Takahashi, M., Fukuda, M., Shibayama, T., Ishikawa, T., Tsuda, F., Tanaka, T., Nishizawa, T. & Okamoto, H. (2000). Visualization of TT virus particles recovered from the sera and feces of infected humans. *Biochemical and Biophysical Research Communications* **279**, 718-724.

Jackson, B., Harvey, C., Galbraith, J., Robertson, M., Warren, K., Holyoake, C., Julian, L. & Varsani, A. (2014a). Clinical beak and feather disease virus infection in wild juvenile eastern rosellas of New Zealand; biosecurity implications for wildlife care facilities. *New Zealand Veterinary Journal* **62**, 297-301.

Jackson, B., Lorenzo, A., Theuerkauf, J., Barnaud, A., Duval, T., Guichard, P., Bloc, H., Baouma, A., Stainton, D., Kraberger, S., Murphy, S., Clark, N., Dillon, C., Knight, T. & Varsani, A. (2014b). Preliminary surveillance for beak and feather disease virus in wild parrots of New Caledonia: Implications of a reservoir species for Ouvea Parakeets. *Emu* **114**, 283-289.

Jackson, B., Varsani, A., Holyoake, C., Jakob-Hoff, R., Robertson, I., McInnes, K., Empson, R., Gray, R., Nakagawa, K. & Warren, K. (2015). Emerging infectious disease or evidence of endemicity? A multi-season study of beak and feather disease virus in wild red-crowned parakeets (Cyanoramphus novaezelandiae). *Archives of Virology* **160**, 2283-2292.

Jeske, H., Lütgemeier, M. & Preiß, W. (2001). DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *The EMBO journal* **20**, 6158-6167.

Jiang, S., Steward, G., Jellison, R., Chu, W. & Choi, S. (2004). Abundance, distribution, and diversity of viruses in alkaline, hypersaline Mono Lake, California. *Microbial Ecology* **47**, 9-17.

Johne, R., Müller, H., Rector, A., van Ranst, M. & Stevens, H. (2009). Rolling-circle amplification of viral DNA genomes using phi29 polymerase. *Trends in Microbiology* **17**, 205-211.

Jovel, J., Preiß, W. & Jeske, H. (2007). Characterization of DNA intermediates of an arising geminivirus. *Virus Research* **130**, 63-70.

Jovel, J., Reski, G., Rothenstein, D., Ringel, M., Frischmuth, T. & Jeske, H. (2004). Sida micrantha mosaic is associated with a complex infection of begomoviruses different from Abutilon mosaic virus. *Archives of Virology* **149**, 829-841.

Julian, L., Lorenzo, A., Chenuet, J. P., Bonzon, M., Marchal, C., Vignon, L., Collings, D. A., Walters, M., Jackson, B. & Varsani, A. (2012). Evidence of multiple introductions of beak and feather disease virus into

the Pacific islands of Nouvelle-Calédonie (New Caledonia). *Journal of General Virology* **93**, 2466-2472.

Julian, L., Piasecki, T., Chrzastek, K., Walters, M., Muhire, B., Harkins, G. W., Martin, D. P. & Varsani, A. (2013). Extensive recombination detected among beak and feather disease virus isolates from breeding facilities in Poland. *Journal of General Virology* **94**, 1086-1095.

Kapoor, A., Dubovi, E. J., Henriquez-Rivera, J. A. & Lipkin, W. I. (2012). Complete genome sequence of the first canine circovirus. *Journal of Virology* **86**, 7018.

Karan, M., Harding, R. M. & Dale, J. L. (1994). Evidence for two groups of banana bunchy top virus isolates. *Journal of General Virology* **75**, 3541-3546.

Kepner, R. L., Jr., Wharton, R. A., Jr. & Suttle, C. A. (1998). Viruses in Antarctic lakes. *Limnology and Oceanography* **43**, 1754-1761.

Kerepesi, C. & Grolmusz, V. (2015). Giant viruses of the Kutch Desert. *Archives of Virology*, 1-4.

Khamrin, P., Okame, M., Thongprachum, A., Nantachit, N., Nishimura, S., Okitsu, S., Maneekarn, N. & Ushijima, H. (2011). A single-tube multiplex PCR for rapid detection in feces of 10 viruses causing diarrhea. *Journal of Virological Methods* **173**, 390-393.

Khan, S. A. (1997). Rolling-circle replication of bacterial plasmids. *Microbiology and Molecular Biology Reviews* **61**, 442-455.

Kim, A. R., Chung, H. C., Kim, H. K., Kim, E. O., Van Nguyen, G., Choi, M. G., Yang, H. J., Kim, J. A. & Park, B. K. (2014). Characterization of a complete genome of a circular single-stranded DNA virus from porcine stools in Korea. *Virus Genes* **48**, 81-88.

Kim, H. K., Park, S. J., van Nguyen, G., Song, D. S., Moon, H. J., Kang, B. K. & Park, B. K. (2012). Identification of a novel single-stranded, circular dna virus from bovine stool. *Journal of General Virology* **93**, 635-639.

Kim, K. H., Chang, H. W., Nam, Y. D., Roh, S. W., Kim, M. S., Sung, Y., Jeon, C. O., Oh, H. M. & Bae, J. W. (2008). Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Applied and Environmental Microbiology* **74**, 5975-5985.

Kim, Y., Aw, T. G., Teal, T. K. & Rose, J. B. (2015). Metagenomic Investigation of Viral Communities in Ballast Water. *Environmental Science and Technology* **49**, 8396-8407.

King, A. M., Lefkowitz, E., Adams, M. J. & Carstens, E. B. (2011). *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*: Burlington: Elsevier Science.

Kirby, A. & Iturriza-Gómara, M. (2012). Norovirus diagnostics: Options, applications and interpretations. *Expert Review of Anti-Infective Therapy* **10**, 423-433.

Koonin, E. V. (1993). A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic Acids Research* **21**, 2541-2547.

Koonin, E. V. & Ilyina, T. V. (1992). Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. *Journal of General Virology* **73**, 2763-2766.

Korlach, J., Bjornson, K. P., Chaudhuri, B. P., Cicero, R. L., Flusberg, B. A., Gray, J. J., Holden, D., Saxena, R., Wegener, J. & Turner, S. W. (2010). Real-time DNA sequencing from single polymerase molecules. *Methods in Enzymology* **472**, 431-455.

Kraberger, S., Argüello-Astorga, G. R., Greenfield, L. G., Galilee, C., Law, D., Martin, D. P. & Varsani, A. (2015a). Characterisation of a diverse range of circular replication-associated protein encoding DNA viruses recovered from a sewage treatment oxidation pond. *Infection, Genetics and Evolution* **31**, 73-86.

Kraberger, S., Farkas, K., Bernardo, P., Booker, C., Argüello-Astorga, G. R., Mesléard, F., Martin, D. P., Roumagnac, P. & Varsani, A. (2015b). Identification of novel Bromus- and Trifolium-associated circular DNA viruses. *Archives of Virology*.

Kraberger, S., Stainton, D., Dayaram, A., Zawar-Reza, P., Gomez, C., Harding, J. S. & Varsani, A. (2013). Discovery of Sclerotinia sclerotiorum Hypovirulence-Associated Virus-1 in Urban River Sediments of Heathcote and Styx Rivers in Christchurch City, New Zealand. *Genome Announcements* **1**, e00559-00513.

Krupovic, M., Zhi, N., Li, J., Hu, G., Koonin, E. V., Wong, S., Shevchenko, S., Zhao, K. & Young, N. S. (2015). Multiple layers of chimerism in a single-stranded DNA virus discovered by deep sequencing. *Genome Biology and Evolution* **7**, 993-1001.

Kumar, J., Gunapati, S., Singh, S. P., Kumar, A., Lalit, A., Sharma, N. C., Puranik, R. & Tuli, R. (2013). A new betasatellite associated with cotton leaf curl Burewala virus infecting tomato in India: Influence on symptoms and viral accumulation. *Archives of Virology* **158**, 1349-1353.

Kumari, S. G., Makkouk, K. M., Loh, M. H., Negassi, K., Tsegay, S., Kidane, R., Kibret, A. & Tesfatsion, Y. (2008). Viral diseases affecting chickpea crops in Eritrea. *Phytopathologia Mediterranea* **47**, 42-49.

Kunik, T., Palanichelvam, K., Czosnek, H., Citovsky, V. & Gafni, Y. (1998). Nuclear import of the capsid protein of tomato yellow leaf curl virus (TYLCV) in plant and insect cells. *Plant Journal* **13**, 393-399.

Labonté, J. M. & Suttle, C. A. (2013). Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME Journal* **7**, 2169-2177.

Lamberto, I., Gunst, K., Müller, H., zur Hausen, H. & de Villiers, E. M. (2014). Mycovirus-like DNA virus sequences from cattle serum and human brain and serum samples from multiple sclerosis patients. *Genome Announcements* **2**, e00848-00814.

Laufs, J., Schumacher, S., Geisler, N., Jupin, I. & Gronenborn, B. (1995a). Identification of the nicking tyrosine of geminivirus Rep protein. *FEBS Letters* **377**, 258-262.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S. G., Schell, J. & Gronenborn, B. (1995b). In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 3879-3883.

Laybourn-Parry, J., Hofer, J. S. & Sommaruga, R. (2001). Viruses in the plankton of freshwater and saline Antarctic lakes. *Freshwater Biology* **46**, 1279-1287.

Le Romancer, M., Gaillard, M., Geslin, C. & Prieur, D. (2007). Viruses in extreme environments. *Reviews in Environmental Science and Biotechnology* **6**, 17-31.

Leary, T. P., Erker, J. C., Chalmers, M. L., Desai, S. M. & Mushahwar, I. K. (1999). Improved detection systems for TT virus reveal high prevalence in humans, non-human primates and farm animals. *Journal of General Virology* **80**, 2115-2120.

Lefeuvre, P., Lett, J. M., Varsani, A. & Martin, D. P. (2009). Widely conserved recombination patterns among single-stranded DNA viruses. *Journal of Virology* **83**, 2697-2707.

Leppik, L., Gunst, K., Lehtinen, M., Dillner, J., Streker, K. & De Villiers, E. M. (2007). In vivo and in vitro intragenomic rearrangement of TT viruses. *Journal of Virology* **81**, 9346-9356.

Levene, H. J., Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G. & Webb, W. W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682-686.

Li, L., Kapoor, A., Slikas, B., Bamidele, O. S., Wang, C., Shaukat, S., Masroor, M. A., Wilson, M. L., Ndjango, J. B. N., Peeters, M., Gross-Camp, N. D., Muller, M. N., Hahn, B. H., Wolfe, N. D., Triki, H., Bartkus, J., Zaidi, S. Z. & Delwart, E. (2010a). Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. *Journal of Virology* **84**, 1674-1682.

Li, L., McGraw, S., Zhu, K., Leutenegger, C. M., Marks, S. L., Kubiski, S., Gaffney, P., Dela Cruz Jr, F. N., Wang, C., Delwart, E. & Pesavento, P. A. (2013). Circovirus in tissues of dogs with vasculitis and hemorrhage. *Emerging Infectious Diseases* **19**, 534-541.

Li, L., Shan, T., Soji, O. B., Alam, M. M., Kunz, T. H., Zaidi, S. Z. & Delwart, E. (2011). Possible cross-species transmission of circoviruses and cycloviruses among farm animals. *Journal of General Virology* **92**, 768-772.

Li, L., Victoria, J. G., Wang, C., Jones, M., Fellers, G. M., Kunz, T. H. & Delwart, E. (2010b). Bat guano virome: Predominance of dietary viruses from insects and plants plus novel mammalian viruses. *Journal of Virology* **84**, 6955-6965.

Li, W., Gu, Y., Shen, Q., Yang, S., Wang, X., Wan, Y. & Zhang, W. (2015). A novel gemycircularvirus from experimental rats. *Virus Genes*.

Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. (2003). Mechanism and regulation of human non-homologous DNA end-joining. *Nature Reviews Molecular Cell Biology* **4**, 712-720.

Lima, F. E., Cibulski, S. P., Dos Santos, H. F., Teixeira, T. F., Varela, A. P., Roehe, P. M., Delwart, E. & Franco, A. C. (2015). Genomic characterization of novel circular ssDNA viruses from insectivorous bats in Southern Brazil. *PLoS One* **10**, e0118070.

Liu, H., Boulton, M. I., Thomas, C. L., Prior, D. A. M., Oparka, K. J. & Davies, J. W. (1999a). Maize streak virus coat protein is karyophyllic and facilitates nuclear transport of viral DNA. *Molecular Plant-Microbe Interactions* **12**, 894-900.

Liu, H., Fu, Y., Li, B., Yu, X., Xie, J., Cheng, J., Ghabrial, S. A., Li, G., Yi, X. & Jiang, D. (2011). Widespread horizontal gene transfer from circular

single-stranded DNA viruses to eukaryotic genomes. *BMC Evolutionary Biology* **11**, 276.

Liu, L., Davies, J. W. & Stanley, J. (1998). Mutational analysis of bean yellow dwarf virus, a geminivirus of the genus Mastrevirus that is adapted to dicotyledonous plants. *Journal of General Virology* **79**, 2265-2274.

Liu, L., Saunders, K., Thomas, C. L., Davies, J. W. & Stanley, J. (1999b). Bean Yellow Dwarf Virus RepA, but Not Rep, Binds to Maize Retinoblastoma Protein, and the Virus Tolerates Mutations in the Consensus Binding Motif. *Virology* **256**, 270-279.

Londoño, A., Riego-Ruiz, L. & Argüello-Astorga, G. R. (2010). DNA-binding specificity determinants of replication proteins encoded by eukaryotic ssDNA viruses are adjacent to widely separated RCR conserved motifs. *Archives of Virology* **155**, 1033-1046.

López-Bueno, A., Tamames, J., Velázquez, D., Moya, A., Quesada, A. & Alcamí, A. (2009). High diversity of the viral community from an Antarctic lake. *Science* **326**, 858-861.

Lorincz, M., Cságola, A., Farkas, S. L., Székely, C. & Tuboly, T. (2011). First detection and analysis of a fish circovirus. *Journal of General Virology* **92**, 1817-1821.

Lyttle, D. J. & Guy, P. L. (2004). First record of Geminiviruses in New Zealand: Abutilon mosaic virus and Honeysuckle yellow vein virus. *Australasian Plant Pathology* **33**, 321-322.

Male, M. F., Kami, V., Kraberger, S. & Varsani, A. (2015). Genome Sequences of Poaceae-Associated Gemycircularviruses from the Pacific Ocean Island of Tonga. *Genome Announcements* **3**.

Male, M. F., Kraberger, S., Stainton, D., Kami, V. & Varsani, A. (2016). Cycloviruses, gemycircularviruses and other novel replication-associated protein encoding circular viruses in Pacific flying fox (Pteropus tonganus) faeces. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*.

Malik, A. H., Briddon, R. W. & Mansoor, S. (2011). Infectious clones of Tomato leaf curl Palampur virus with a defective DNA B and their pseudo-recombination with Tomato leaf curl New Delhi virus. *Virology Journal* **8**.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F. & Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380.

Martin, D. P., Biagini, P., Lefeuvre, P., Golden, M., Roumagnac, P. & Varsani, A. (2011a). Recombination in eukaryotic single stranded DNA viruses. *Viruses* **3**, 1699-1738.

Martin, D. P., Briddon, R. W. & Varsani, A. (2011b). Recombination patterns in dicot-infecting mastreviruses mirror those found in monocot-infecting mastreviruses. *Archives of Virology* **156**, 1463-1469.

Marzano, S. L. & Domier, L. (2015). Novel mycoviruses discovered from metatranscriptomics survey of soybean phyllosphere phytobiomes. *Virus Research*.

Massaro, M., Ortiz-Catedral, L., Julian, L., Galbraith, J. A., Kurenbach, B., Kearvell, J., Kemp, J., van Hal, J., Elkington, S., Taylor, G., Greene, T., van de Wetering, J., van de Wetering, M., Pryde, M., Dilks, P., Heber, S., Steeves, T. E., Walters, M., Shaw, S., Potter, J., Farrant, M., Brunton, D. H., Hauber, M., Jackson, B., Bell, P., Moorhouse, R., McInnes, K. & Varsani, A. (2012). Molecular characterisation of beak and feather disease virus (BFDV) in New Zealand and its implications for managing an infectious disease. *Archives of Virology* **157**, 1651-1663.

Mauricio-Castillo, J. A., Torres-Herrera, S. I., Cárdenas-Conejo, Y., Pastor-Palacios, G., Méndez-Lozano, J. & Argüello-Astorga, G. R. (2014). A novel begomovirus isolated from sida contains putative cis- and trans-acting replication specificity determinants that have evolved independently in several geographical lineages. *Archives of virology* **159**, 2283-2294.

McDaniel, L. D., Rosario, K., Breitbart, M. & Paul, J. H. (2014). Comparative metagenomics: Natural populations of induced prophages demonstrate highly unique, lower diversity viral sequences. *Environmental Microbiology* **16**, 570-585.

Merchant, C. A., Healy, K., Wanunu, M., Ray, V., Peterman, N., Bartel, J., Fischbein, M. D., Venta, K., Luo, Z., Johnson, A. T. C. & Drndić, M. (2010). DNA translocation through graphene nanopores. *Nano Letters* **10**, 2915-2921.

Merriman, B., Torrent, I. & Rothberg, J. M. (2012). Progress in Ion Torrent semiconductor chip based sequencing. *Electrophoresis* **33**, 3397-3417.

Metzker, M. L. (2010). Sequencing technologies the next generation. *Nature Reviews Genetics* **11**, 31-46.

Mochizuki, M., Ohshima, T., Une, Y. & Yachi, A. (2008). Recombination between vaccine and field strains of canine parvovirus is revealed by isolation of virus in canine and feline cell cultures. *Journal of Veterinary Medical Science* **70**, 1305-1314.

Mochizuki, T., Krupovic, M., Pehau-Arnaudet, G., Sako, Y., Forterre, P. & Prangishvili, D. (2012). Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 13386-13391.

Muhire, B. M., Varsani, A. & Martin, D. P. (2014). SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE* **9**, e108277.

Mushahwar, I. K., Erker, J. C., Muerhoff, A. S., Leary, T. P., Simons, J. N., Birkenmeyer, L. G., Chalmers, M. L., Pilot-Matias, T. J. & Dexai, S. M. (1999). Molecular and biophysical characterization of TT virus: Evidence for a new virus family infecting humans. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 3177-3182.

Nakamura, S., Yang, C. S., Sakon, N., Ueda, M., Tougan, T., Yamashita, A., Goto, N., Takahashi, K., Yasunaga, T., Ikuta, K., Mizutani, T., Okamoto, Y., Tagami, M., Morita, R., Maeda, N., Kawai, J., Hayashizaki, Y., Nagai, Y., Horii, T., Iida, T. & Nakaya, T. (2009). Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS ONE* **4**.

Nash, T. E., Dallas, M. B., Reyes, M. I., Buhrman, G. K., Ascencio-Ibañez, J. T. & Hanley-Bowdoin, L. (2011). Functional analysis of a novel motif conserved across geminivirus Rep proteins. *Journal of Virology* **85**, 1182-1192.

Nelson, J. R., Cai, Y. C., Giesler, T. L., Farchaus, J. W., Sundaram, S. T., Ortiz-Rivera, M., Hosta, L. P., Hewitt, P. L., Mamone, J. A., Palaniappan, C. & Fuller, C. W. (2002). TempliPhi, φ29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. *BioTechniques* **32**, S44-S47.

Neumann, E. J., Dobbinson, S. S., Welch, E. B. & Morris, R. S. (2007). Descriptive summary of an outbreak of porcine post-weaning multisystemic wasting syndrome (PMWS) in New Zealand. *New Zealand Veterinary Journal* **55**, 346-352.

Ng, T. F. F., Alavandi, S., Varsani, A., Burghart, S. & Breitbart, M. (2013). Metagenomic identification of a nodavirus and a circular ssDNA virus in semi-purified viral nucleic acids from the hepatopancreas of healthy Farfantepenaeus duorarum shrimp. *Diseases of Aquatic Organisms* **105**, 237-242.

Ng, T. F. F., Chen, L. F., Zhou, Y., Shapiro, B., Stiller, M., Heintzman, P. D., Varsani, A., Kondov, N. O., Wong, W., Deng, X., Andrews, T. D., Moorman, B. J., Meulendyk, T., Mackay, G., Gilbertson, R. L., Delwart, E. & Palese, P. (2014). Preservation of viral genomes in 700-y-old caribou feces from a subarctic ice patch. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 16842-16847.

Ng, T. F. F., Duffy, S., Polston, J. E., Bixby, E., Vallad, G. E. & Breitbart, M. (2011a). Exploring the diversity of plant DNA viruses and their satellites using vector-enabled metagenomics on whiteflies. *PLoS ONE* **6**.

Ng, T. F. F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B. S., Eric Wommack, K. & Delwarta, E. (2012). High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *Journal of Virology* **86**, 12161-12175.

Ng, T. F. F., Willner, D. L., Lim, Y. W., Schmieder, R., Chau, B., Nilsson, C., Anthony, S., Ruan, Y., Rohwer, F. & Breitbart, M. (2011b). Broad surveys of DNA viral diversity obtained through viral metagenomics of mosquitoes. *PLoS ONE* **6**, e20579.

Ng, T. F. F., Zhang, W., Sachsenröder, J., Kondov, N. O., da Costa, A. C., Vega, E., Holtz, L. R., Wu, G., Wang, D., Stine, C. O., Antonio, M., Mulvaney, U. S., Muench, M. O., Deng, X., Ambert-Balay, K., Pothier, P., Vinjé, J. & Delwart, E. (2015). A diverse group of small circular ssDNA viral genomes in human and non-human primate stools. *Virus Evolution* **1**, vev017.

Nguyen, V. G., Kim, H. K., Moon, H. J., Park, S. J., Keum, H. O., Rho, S., Han, J. Y. & Park, B. K. (2012). Population dynamics and ORF3 gene

evolution of porcine circovirus type 2 circulating in Korea. *Archives of Virology* **157**, 799-810.

Nishiyama, S., Dutia, B. M., Stewart, J. P., Meredith, A. L., Shaw, D. J., Simmonds, P. & Sharp, C. P. (2014). Identification of novel anelloviruses with broad diversity in UK rodents. *Journal of General Virology* **95**, 1544-1553.

Okamoto, H., Takahashi, M., Nishizawa, T., Tawara, A., Fukai, K., Muramatsu, U., Naito, Y. & Yoshikawa, A. (2002). Genomic characterization of TT viruses (TTVs) in pigs, cats and dogs and their relatedness with species-specific TTVs in primates and tupaias. *Journal of General Virology* **83**, 1291-1297.

Oren, A., Bratbak, G. & Heldal, M. (1997). Occurrence of virus-like particles in the Dead Sea. *Extremophiles* **1**, 143-149.

Orozco, B. M. & Hanley-Bowdoin, L. (1998). Conserved sequence and structural motifs contribute to the DNA binding and cleavage activities of a geminivirus replication protein. *Journal of Biological Chemistry* **273**, 24448-24456.

Ortiz-Catedral, L., Kurenbach, B., Massaro, M., McInnes, K., Brunton, D. H., Hauber, M. E., Martin, D. P. & Varsani, A. (2010). A new isolate of beak and feather disease virus from endemic wild red-fronted parakeets (Cyanoramphus novaezelandiae) in New Zealand. *Archives of Virology* **155**, 613-620.

Ortmann, A. C. & Suttle, C. A. (2005). High abundances of viruses in a deep-sea hydrothermal vent system indicates viral mediated microbial mortality. *Deep-Sea Research Part I: Oceanographic Research Papers* **52**, 1515-1527.

Padilla-Rodriguez, M., Rosario, K. & Breitbart, M. (2013). Novel cyclovirus discovered in the Florida woods cockroach *Eurycotis floridana* (Walker). *Archives of Virology* **158**, 1389-1392.

Paprotka, T., Deuschle, K., Pilartz, M. & Jeske, H. (2015). Form follows function in geminiviral minichromosome architecture. *Virus Research* **196**, 44-55.

Pathrose, B., Jones, E. E., Jaspers, M. V. & Ridgway, H. J. (2014). High genotypic and virulence diversity in Ilyonectria liriodendri isolates associated with black foot disease in New Zealand vineyards. *Plant Pathology* **63**, 613-624.

Patil, B. L. & Fauquet, C. M. (2010). Differential interaction between cassava mosaic geminiviruses and geminivirus satellites. *Journal of General Virology* **91**, 1871-1882.

Pham, H. T., Bergoin, M. & Tijssen, P. (2013). Acheta domesticus Volvovirus, a Novel Single-Stranded Circular DNA Virus of the House Cricket. *Genome Announcements* **1**, 00079-00013.

Phan, T. G., da Costa, A. C., Del Valle Mendoza, J., Bucardo-Rivera, F., Nordgren, J., O'Ryan, M., Deng, X. & Delwart, E. (2016). The fecal virome of South and Central American children with diarrhea includes small circular DNA viral genomes of unknown origin. *Archives of Virology*, 1-8.

Phan, T. G., Kapusinszky, B., Wang, C., Rose, R. K., Lipton, H. L. & Delwart, E. L. (2011). The fecal viral flora of wild rodents. *PLoS Pathogens* **7**.

Phan, T. G., Mori, D., Deng, X., Rajidrajith, S., Ranawaka, U., Ng, T. F. F. & Delwart, E. (2015). Small circular single stranded DNA viral genomes in unexplained cases of human encephalitis, diarrhea, and in untreated sewage. *Virology* **482**, 98-104.

Phan, T. G., Vo, N. P., Boros, Á., Pankovics, P., Reuter, G., Li, O. T. W., Wang, C., Deng, X., Poon, L. L. M. & Delwart, E. (2013). The Viruses of Wild Pigeon Droppings. *PLoS ONE* **8**.

Pie, J., Kim, B. & Grishin, N. (2008). PROMALS3D: a tool for multiple sequence and structure alignment. *Nucleic Acids Research* **36**, 2295-2300.

Pilartz, M. & Jeske, H. (1992). Abutilon mosaic geminivirus double-stranded DNA is packed into minichromosomes. *Virology* **189**, 800-802.

Prestel, E., Regeard, C., Salamitou, S., Neveu, J. & Dubow, M. S. (2013). The bacteria and bacteriophages from a Mesquite Flats site of the Death Valley desert. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* **103**, 1329-1341.

Prigent, M., Leroy, M., Confalonieri, F., Dutertre, M. & DuBow, M. S. (2005). A diversity of bacteriophage forms and genomes can be isolated from the surface sands of the Sahara Desert. *Extremophiles* **9**, 289-296.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P. & Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**.

Rachel, R., Bettstetter, M., Hedlund, B. P., Häring, M., Kessler, A., Stetter, K. O. & Prangishvili, D. (2002). Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. *Archives of Virology* **147**, 2419-2429.

Ramos, A. P. D., Stefanelli, C. C., Linhares, R. E. C., Brito, B. G. D., Santos, N., Gouvea, V., Lima, R. D. C. & Nozawa, C. (2000). The stability of porcine rotavirus in feces. *Veterinary Microbiology* **71**, 1-8.

Reavy, B., Swanson, M. M., Cock, P. J. A., Dawson, L., Freitag, T. E., Singh, B. K., Torrance, L., Mushegian, A. R. & Taliansky, M. (2015). Distinct circular single-stranded DNA viruses exist in different soil types. *Applied and Environmental Microbiology* **81**, 3934-3945.

Reuter, G., Boros, Á., Delwart, E. & Pankovics, P. (2014). Novel circular single-stranded DNA virus from turkey faeces. *Archives of Virology* **159**, 2161-2164.

Ritchie, B. W., Niagro, F. D., Lukert, P. D., Steffens Iii, W. L. & Latimer, K. S. (1989). Characterization of a new virus from cockatoos with psittacine beak and feather disease. *Virology* **171**, 83-88.

Ritchie, P. A., Anderson, I. L. & Lambert, D. M. (2003). Evidence for specificity of psittacine beak and feather disease viruses among avian hosts. *Virology* **306**, 109-115.

Rohde, W., Randles, J. W., Langridge, P. & Hanold, D. (1990). Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus. *Virology* **176**, 648-651.

Roine, E., Kukkaro, P., Paulin, L., Laurinavičius, S., Domanska, A., Somerharju, P. & Bamford, D. H. (2010). New, closely related haloarchaeal viral elements with different nucleic acid types. *Journal of Virology* **84**, 3682-3689.

Rokyta, D. R., Burch, C. L., Caudle, S. B. & Wichman, H. A. (2006). Horizontal gene transfer and the evolution of microvirid coliphage genomes. *Journal of Bacteriology* **188**, 1134-1142.

Roossinck, M. J. (1997). Mechanisms of plant virus evolution. In *Annual Review of Phytopathology*, pp. 191-209.

Rosario, K., Dayaram, A., Marinov, M., Ware, J., Kraberger, S., Stainton, D., Breitbart, M. & Varsani, A. (2012a). Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). *Journal of General Virology* **93**, 2668-2681.

Rosario, K., Duffy, S. & Breitbart, M. (2009a). Diverse circovirus-like genome architectures revealed by environmental metagenomics. *Journal of General Virology* **90**, 2418-2424.

Rosario, K., Duffy, S. & Breitbart, M. (2012b). A field guide to eukaryotic circular single-stranded DNA viruses: Insights gained from metagenomics. *Archives of Virology* **157**, 1851-1871.

Rosario, K., Marinov, M., Stainton, D., Kraberger, S., Wiltshire, E. J., Collings, D. A., Walters, M., Martin, D. P., Breitbart, M. & Varsani, A. (2011). Dragonfly cyclovirus, a novel single-stranded DNA virus discovered in dragonflies (Odonata: Anisoptera). *Journal of General Virology* **92**, 1302-1308.

Rosario, K., Nilsson, C., Lim, Y. W., Ruan, Y. & Breitbart, M. (2009b). Metagenomic analysis of viruses in reclaimed water. *Environmental Microbiology* **11**, 2806-2820.

Rosario, K., Schenck, R. O., Harbeitner, R. C., Lawler, S. N. & Breitbart, M. (2015a). Novel circular single-stranded DNA viruses identified in marine invertebrates reveal high sequence diversity and consistent predicted intrinsic disorder patterns within putative structural proteins. *Frontiers in Microbiology* **6**, 696.

Rosario, K., Seah, Y. M., Marr, C., Varsani, A., Kraberger, S., Stainton, D., Moriones, E., Polston, J. E., Duffy, S. & Breitbart, M. (2015b). Vector-enabled metagenomic (VEM) surveys using whiteflies (Aleyrodidae) reveal novel begomovirus species in the new and old worlds. *Viruses* **7**, 5553-5570.

Rothberg, J. M. & Leamon, J. H. (2008). The development and impact of 454 sequencing. *Nature Biotechnology* **26**, 1117-1124.

Roumagnac, P., Granier, M., Bernardo, P., Deshoux, M., Ferdinand, R., Galzi, S., Fernandez, E., Julian, C., Abt, I., Filloux, D., Mesléard, F., Varsani, A., Blanc, S., Martin, D. P. & Peterschmitt, M. (2015). Alfalfa leaf curl virus: An aphid-transmitted geminivirus. *Journal of Virology* **89**, 9683-9688.

Roux, S., Enault, F., Bronner, G., Vaulot, D., Forterre, P. & Krupovic, M. (2013). Chimeric viruses blur the borders between the major groups of eukaryotic single-stranded DNA viruses. *Nature Communications* **4**.

Roux, S., Enault, F., Robin, A., Ravet, V., Personnic, S., Theil, S., Colombet, J., Sime-Ngando, T. & Debroas, D. (2012). Assessing the diversity and specificity of two freshwater viral communities through metagenomics. *PLoS ONE* **7**.

Sachsenröder, J., Braun, A., Machnowska, P., Ng, T. F. F., Deng, X., Guenther, S., Bernstein, S., Ulrich, R. G., Delwart, E. & Johne, R. (2014). Metagenomic identification of novel enteric viruses in urban wild

rats and genome characterization of a group A rotavirus. *Journal of General Virology* **95**, 2734-2747.

Sachsenröder, J., Twardziok, S., Hammerl, J. A., Janczyk, P., Wrede, P., Hertwig, S. & Johne, R. (2012). Simultaneous identification of DNA and RNA viruses present in pig faeces using process-controlled deep sequencing. *PLoS ONE* **7**, e0034631.

Santos, F., Moreno-Paz, M., Meseguer, I., López, C., Rosselló-Mora, R., Parro, V. & Antón, J. (2011). Metatranscriptomic analysis of extremely halophilic viral communities. *ISME Journal* **5**, 1621-1633.

Sasaki, M., Orba, Y., Ueno, K., Ishii, A., Moonga, L., Hangombe, B. M., Mweene, A. S., Ito, K. & Sawa, H. (2015). Metagenomic analysis of the shrew enteric virome reveals novel viruses related to human stool-associated viruses. *Journal of General Virology* **96**, 440-452.

Sato, G., Kawashima, T., Kiuchi, M. & Tohya, Y. (2015). Novel cyclovirus detected in the intestinal contents of Taiwan squirrels (Callosciurus erythraeus thaiwanensis). *Virus Genes* **51**, 148-151.

Saunders, K., Bedford, I. D., Briddon, R. W., Markham, P. G., Wong, S. M. & Stanley, J. (2000). A unique virus complex causes Ageratum yellow vein disease. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6890-6895.

Saunders, K., Lucy, A. & Stanley, J. (1991). DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. *Nucleic Acids Research* **19**, 2325-2330.

Saunders, K., Salim, N., Mali, V. R., Malathi, V. G., Briddon, R., Markham, P. G. & Stanley, J. (2002). Characterisation of Sri Lankan cassava mosaic virus and Indian cassava mosaic virus: Evidence for acquisition of a DNA B component by a monopartite begomovirus. *Virology* **293**, 63-74.

Saunders, K. & Stanley, J. (1999). A nanovirus-like DNA component associated with yellow vein disease of Ageratum conyzoides: Evidence for interfamilial recombination between plant DNA viruses. *Virology* **264**, 142-152.

Sauvage, V., Cheval, J., Foulongne, V., Gouilh, M. A., Pariente, K., Manuguerra, J. C., Richardson, J., Dereure, O., Lecuit, M., Burguiere, A., Caro, V. & Eloit, M. (2011). Identification of the first human Gyrovirus, a virus related to chicken anemia virus. *Journal of Virology* **85**, 7948-7950.

Savory, F. R. & Ramakrishnan, U. (2014). Asymmetric patterns of reassortment and concerted evolution in Cardamom bushy dwarf virus. *Infection, Genetics and Evolution* **24**, 15-24.

Schadt, E. E., Turner, S. & Kasarskis, A. (2010). A window into third-generation sequencing. *Human Molecular Genetics* **19**, R227-R240.

Schneider, G. F., Kowalczyk, S. W., Calado, V. E., Pandraud, G., Zandbergen, H. W., Vandersypen, L. M. K. & Dekker, C. (2010). DNA translocation through graphene nanopores. *Nano Letters* **10**, 3163-3167.

Shan, T., Li, L., Simmonds, P., Wang, C., Moeser, A. & Delwart, E. (2011). The fecal virome of pigs on a high-density farm. *Journal of Virology* **85**, 11697-11708.

Shendure, J. & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology* **26**, 1135-1145.

Siegl, G., Hallauer, C., Novak, A. & Kronauer, G. (1971). Parvoviruses as contaminants of permanent human cell lines - II. Physicochemical properties of the isolated viruses. *Archiv für die gesamte Virusforschung* **35**, 91-103.

Sikorski, A., Argüello-Astorga, G. R., Dayaram, A., Dobson, R. C. J. & Varsani, A. (2013a). Discovery of a novel circular single-stranded DNA virus from porcine faeces. *Archives of Virology* **158**, 283-289.

Sikorski, A., Dayaram, A. & Varsani, A. (2013b). Identification of a Novel Circular DNA Virus in New Zealand Fur Seal (Arctocephalus forsteri) Fecal Matter. *Genome Announcements* **1**, 00558-00513.

Sikorski, A., Kearvell, J., Elkington, S., Dayaram, A., Argüello-Astorga, G. R. & Varsani, A. (2013c). Novel ssDNA viruses discovered in yellow-crowned parakeet (Cyanoramphus auriceps) nesting material. *Archives of Virology* **158**, 1603-1607.

Sikorski, A., Massaro, M., Kraberger, S., Young, L. M., Smalley, D., Martin, D. P. & Varsani, A. (2013d). Novel myco-like DNA viruses discovered in the faecal matter of various animals. *Virus Research* **177**, 209-216.

Silva, F. N., Lima, A. T., Rocha, C. S., Castillo-Urquiza, G. P., Alves-Júnior, M. & Zerbini, F. M. (2014). Recombination and pseudorecombination driving the evolution of the begomoviruses Tomato severe rugose virus (ToSRV) and Tomato rugose mosaic virus (ToRMV): Two recombinant DNA-A components sharing the same DNA-B. *Virology Journal* **11**.

Sime-Ngando, T., Lucas, S., Robin, A., Tucker, K. P., Colombet, J., Bettarel, Y., Desmond, E., Gribaldo, S., Forterre, P., Breitbart, M. & Prangishvili, D. (2011). Diversity of virus-host systems in hypersaline Lake Retba, Senegal. *Environmental Microbiology* **13**, 1956-1972.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. & Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Research* **19**, 1117-1123.

Smith, R. J., Jeffries, T. C., Roudnew, B., Seymour, J. R., Fitch, A. J., Simons, K. L., Speck, P. G., Newton, K., Brown, M. H. & Mitchell, J. G. (2013). Confined aquifers as viral reservoirs. *Environmental Microbiology Reports* **5**, 725-730.

Smits, S. L., Schapendonk, C. M. E., van Beek, J., Vennema, H., Schürch, A. C., Schipper, D., Bodewes, R., Haagmans, B. L., Osterhaus, A. D. M. E. & Koopmans, M. P. (2014). New viruses in idiopathic human diarrhea cases, the Netherlands. *Emerging Infectious Diseases* **20**, 1218-1222.

Soffer, N., Brandt, M. E., Correa, A. M., Smith, T. B. & Thurber, R. V. (2014). Potential role of viruses in white plague coral disease. *ISME J* **8**, 271-283.

Spiers, A. G. (1998). Melampsora and Marssonina pathogens of poplars and willows in New Zealand. *European Journal of Forest Pathology* **28**, 233-240.

Stainton, D., Kraberger, S., Walters, M., Wiltshire, E. J., Rosario, K., Halafihi, M., Lolohea, S., Katoa, I., Faitua, T. H., Aholelei, W., Taufa, L., Thomas, J. E., Collings, D. A., Martin, D. P. & Varsani, A. (2012). Evidence of inter-component recombination, intra-component recombination and reassortment in banana bunchy top virus. *Journal of General Virology* **93**, 1103-1119.

Stainton, D., Martin, D. P., Collings, D. A., Thomas, J. E. & Varsani, A. (2016). Identification and in silico characterisation of defective molecules associated with isolates of banana bunchy top virus. *Archives of Virology*, 1-8.

Stainton, D., Martin, D. P., Muhire, B. M., Lolohea, S., Halafihi, M., Lepoint, P., Blomme, B., Crew, K. S., Sharman, M., Kraberger, S., Dayaram, A., Walters, M., Collings, D. A., Mabvakure, B., Lemey, P., Harkins, G. W., Thomas, J. E. & Varsani, A. (2015). The global distributions of Banana bunchy top virus reveals little evidence for frequent recent, human-mediated long distance dispersal events. *Virus Evolution* **1**, 1-16.

Stanley, J. (1983). Infectivity of the cloned geminivirus genome requires sequences from both DNAs. *Nature* **305**, 643-645.

Stanley, J. (2004). Subviral DNAs associated with geminivirus disease complexes. *Veterinary Microbiology* **98**, 121-129.

Stedman, K. (2013). Mechanisms for RNA capture by ssDNA viruses: Grand theft RNA. *Journal of Molecular Evolution* **76**, 359-364.

Steinel, A., Parrish, C. R., Bloom, M. E. & Truyen, U. (2001). Parvovirus infections in wild carnivores. *Journal of Wildlife Diseases* **37**, 594-607.

Steinfeldt, T., Finsterbusch, T. & Mankertz, A. (2006). Demonstration of nicking/joining activity at the origin of DNA replication associated with the Rep and Rep' proteins of porcine circovirus type 1. *Journal of Virology* **80**, 6225-6234.

Steinfeldt, T., Finsterbusch, T. & Mankertz, A. (2007). Functional analysis of cis- and trans-acting replication factors of porcine circovirus type 1. *Journal of Virology* **81**, 5696-5704.

Stenger, D. C., Revington, G. N., Stevenson, M. C. & Bisaro, D. M. (1991). Replicational release of geminivirus genomes from tandemly repeated copies: Evidence for rolling-circle replication of a plant viral DNA. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 8029-8033.

Stöver, B. C. & Müller, K. F. (2010). TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* **11**, 7.

Streck, A., Bonatto, S. L., Homeier, T., Souza, C. K., Gonçalves, K. R., Gava, D., Canal, C. W. & Truyen, U. (2011). High rate of viral evolution in the capsid protein of porcine parvovirus. *Journal of General Virology* **92**, 2628-2636.

Sung, Y. K. & Coutts, R. H. A. (1995). Pseudorecombination and complementation between potato yellow mosaic geminivirus and tomato golden mosaic geminivirus. *Journal of General Virology* **76**, 2809-2815.

Tan, L. V., van Doorn, H. R., Nghia, H. D. T., Chau, T. T. H., Tu, L. T. P., de Vries, M., Canuti, M., Deijs, M., Jebbink, M. F., Baker, S., Bryant, J. E., Tham, N. T., Bkrong, N. T. T. C., Boni, M. F., Loi, T. Q., Phuong, L. T., Verhoeven, J. T. P., Crusat, M., Jeeninga, R. E., Schultsz, C., Chau, N. V. V., Hien, T. T., van der Hoek, L., Farrar, J. & de Jong, M. D. (2013). Identification of a new cyclovirus in cerebrospinal fluid of patients with acute central nervous system infections. *mBio* **4**.

Tarján, Z., Pénzes, J., Tóth, R. & Benko, M. (2014). First detection of circovirus-like sequences in amphibians and novel putative circoviruses in fishes. *Acta Veterinaria Hungarica* **62**, 134-144.

Timchenko, T., De Kouchkovsky, F., Katul, L., David, C., Vetten, H. J. & Gronenborn, B. (1999). A single Rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants. *Journal of Virology* **73**, 10173-10182.

Tischer, I., Gelderblom, H., Vettermann, W. & Koch, M. A. (1982). A very small porcine virus with circular single-stranded DNA. *Nature* **295**, 64-66.

Todd, D., Creelan, J. L., Mackie, D. P., Rixon, F. & McNulty, M. S. (1990). Purification and biochemical characterization of chicken anaemia agent. *Journal of General Virology* **71**, 819-823.

Tyumentsev, A. I., Tikunova, N. V., Tikunov, A. Y. & Babkin, I. V. (2014). Recombination in the evolution of human bocavirus. *Infection, Genetics and Evolution* **28**, 11-14.

Uch, R., Fournier, P. E., Robert, C., Blanc-Tailleur, C., Galicher, V., Barre, R., Jordier, F., de Micco, P., Raoult, D. & Biagini, P. (2015). Divergent gemycircularvirus in HIV-positive blood, France. *Emerging Infectious Diseases* **21**, 2096-2098.

Unseld, S., Ringel, M., Höfer, P., Höhnle, M., Jeske, H., Bedford, I. D., Markham, P. G. & Frischmuth, T. (2000). Host range and symptom variation of pseudorecombinant virus produced by two distinct bipartite geminiviruses. *Archives of Virology* **145**, 1449-1454.

van den Brand, J. M. A., van Leeuwen, L., Schapendonk, C. M., Simon, J. H., Haagmans, B. L., Osterhaus, A. D. M. E. & Smits, S. L. (2012). Metagenomic analysis of the viral flora of pine marten and european badger feces. *Journal of Virology* **86**, 2360-2365.

van der Walt, E., Martin, D. P., Varsani, A., Polston, J. E. & Rybicki, E. P. (2008). Experimental observations of rapid maize streak virus evolution reveal a strand-specific nucleotide substitution bias. *Virology Journal* **5**.

van der Walt, E., Rybicki, E. P., Varsani, A., Polston, J. E., Billharz, R., Donaldson, L., Monjane, A. L. & Martin, D. P. (2009). Rapid host adaptation by extensive recombination. *Journal of General Virology* **90**, 734-746.

Varma, A. & Malathi, V. G. (2003). Emerging geminivirus problems: A serious threat to crop production. *Annals of Applied Biology* **142**, 145-164.

Varsani, A., Kraberger, S., Jennings, S., Porzig, E. L., Julian, L., Massaro, M., Pollard, A., Ballard, G. & Ainley, D. G. (2014a). A novel papillomavirus in Adélie penguin (Pygoscelis adeliae) faeces sampled at the Cape Crozier colony, Antarctica. *Journal of General Virology* **95**, 1352-1365.

Varsani, A., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Brown, J. K., Murilo Zerbini, F. & Martin, D. P. (2014b). Establishment of three new genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. *Archives of Virology* **159**, 2193-2203.

Varsani, A., Porzig, E. L., Jennings, S., Kraberger, S., Farkas, K., Julian, L., Masaro, M., Ballard, G. & Ainley, D. G. (2015). Identification of an avian polyomavirus associated with Adelie penguins (Pygoscelis adeliae). *Journal of General Virology* **96**, 851-857.

Varsani, A., Shepherd, D. N., Monjane, A. L., Owor, B. E., Erdmann, J. B., Rybicki, E. P., Peterschmitt, M., Briddon, R. W., Markham, P. G., Oluwafemi, S., Windram, O. P., Lefeuvre, P., Lett, J. M. & Martin, D.

P. (2008). Recombination, decreased host specificity and increased mobility may have driven the emergence of maize streak virus as an agricultural pathogen. *Journal of General Virology* **89**, 2063-2074.

Vega-Rocha, S., Byeon, I. J. L., Gronenborn, B., Gronenborn, A. M. & Campos-Olivas, R. (2007a). Solution Structure, Divalent Metal and DNA Binding of the Endonuclease Domain from the Replication Initiation Protein from Porcine Circovirus 2. *Journal of Molecular Biology* **367**, 473-487.

Vega-Rocha, S., Gronenborn, B., Gronenborn, A. M. & Campos-Olivas, R. (2007b). Solution structure of the endonuclease domain from the master replication initiator protein of the nanovirus faba bean necrotic yellows virus and comparison with the corresponding geminivirus and circoviras structures. *Biochemistry* **46**, 6201-6212.

Venkatesan, B. M. & Bashir, R. (2011). Nanopore sensors for nucleic acid analysis. *Nature Nanotechnology* **6**, 615-624.

Victoria, J. G., Kapoor, A., Li, L., Blinkova, O., Slikas, B., Wang, C., Naeem, A., Zaidi, S. & Delwart, E. (2009). Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *Journal of Virology* **83**, 4642-4651.

Viguera, E., Canceill, D. & Ehrlich, S. D. (2001). Replication slippage involves DNA polymerase pausing and dissociation. *EMBO Journal* **20**, 2587-2595.

Watt, M., Bulman, L. & Palmer, D. (2011). The economic cost of Dothistroma needle blight to the New Zealand forest industry. *New Zealand Journal of Forestry* **56**, 20-22.

Whon, T. W., Kim, M. S., Roh, S. W., Shin, N. R., Lee, H. W. & Bae, J. W. (2012). Metagenomic characterization of airborne viral dna diversity in the near-surface atmosphere. *Journal of Virology* **86**, 8221-8231.

Williamson, S. J., Cary, S. C., Williamson, K. E., Helton, R. R., Bench, S. R., Winget, D. & Wommack, K. E. (2008). Lysogenic virus-host interactions predominate at deep-sea diffuse-flow hydrothermal vents. *ISME Journal* **2**, 1112-1121.

Woo, P. C. Y., Lau, S. K. P., Teng, J. L. L., Tsang, A. K. L., Joseph, M., Wong, E. Y. M., Tang, Y., Sivakumar, S., Bai, R., Wernery, R., Wernery, U. & Yuen, K. Y. (2014). Metagenomic analysis of viromes of dromedary camel fecal samples reveals large number and high diversity of circoviruses and picobirnaviruses. *Virology* **471-473**, 117-125.

Wright, E. A., Heckel, T., Groenendijk, J., Davies, J. W. & Boulton, M. I. (1997). Splicing features in maize streak virus virion- and complementary-sense gene expression. *Plant Journal* **12**, 1285-1297.

Wu, Z., Yang, L., Ren, X., He, G., Zhang, J., Yang, J., Qian, Z., Dong, J., Sun, L., Zhu, Y., Du, J., Yang, F., Zhang, S. & Jin, Q. (2015). Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. *The ISME Journal*.

Xu, Y. & Price, B. D. (2011). Chromatin dynamics and the repair of DNA double strand breaks. *Cell Cycle* **10**, 261-267.

Yamada, T., Kawasaki, T., Nagata, S., Fujiwara, A., Usami, S. & Fujie, M. (2007). New bacteriophages that infect the phytopathogen Ralstonia solanacearum. *Microbiology* **153**, 2630-2639.

Yoon-Robarts, M., Blouin, A. G., Bleker, S., Kleinschmidt, J. A., Aggarwal, A. K., Escalante, C. R. & Linden, R. M. (2004). Residues within the B′ motif are critical for DNA binding by the superfamily 3 helicase Rep40 of adeno-associated virus type 2. *Journal of Biological Chemistry* **279**, 50472-50481.

Yoshida, M., Takaki, Y., Eitoku, M., Nunoura, T. & Takai, K. (2013). Metagenomic analysis of viral communities in (hado)pelagic sediments. *PLoS One* **8**, e57271.

Yu, X., Li, B., Fu, Y., Xie, J., Cheng, J., Ghabrial, S. A., Li, G., Yi, X. & Jiang, D. (2013). Extracellular transmission of a DNA mycovirus and its use as a natural fungicide. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1452-1457.

Yu, X., Li, B., Fu, Y. P., Jiang, D. H., Ghabrial, S. A., Li, G. Q., Peng, Y. L., Xie, J. T., Cheng, J. S., Huang, J. B. & Yi, X. H. (2010). A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 8387-8392.

Zablocki, O., van Zyl, L., Adriaenssens, E. M., Rubagotti, E., Tuffin, M., Cary, S. C. & Cowan, D. (2014). High-level diversity of tailed phages, eukaryote-associated viruses, and virophage-like elements in the metaviromes of antarctic soils. *Applied and Environmental Microbiology* **80**, 6888-6897.

Zawar-Reza, P., Argüello-Astorga, G. R., Kraberger, S., Julian, L., Stainton, D., Broady, P. A. & Varsani, A. (2014). Diverse small circular single-stranded DNA viruses identified in a freshwater pond on the McMurdo Ice Shelf (Antarctica). *Infection, Genetics and Evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* **26**, 132-138.

Zhang, W., Li, L., Deng, X., Kapusinszky, B., Pesavento, P. A. & Delwart, E. (2014). Faecal virome of cats in an animal shelter. *Journal of General Virology* **95**, 2553-2564.

Zhang, W., Olson, N. H., Baker, T. S., Faulkner, L., Agbandje-McKenna, M., Boulton, M. I., Davies, J. W. & McKenna, R. (2001). Structure of the maize streak virus geminate particle. *Virology* **279**, 471-477.